ENERGETICS OF E-BOX RECOGNITION BY TUMOR SUPPRESSOR MAX p21

DISSERTATION

zur Erlangung der naturwissenschaftlichen Doktorwürde (Dr.sc.nat.)

vorgelegt der
Mathematisch-naturwissenschaftlichen Fakultät der Universität Zürich

von

Laura E. Meier-Andrejszki
aus Zürich

Promotionskomitee
Prof. Dr. Markus Grütter
PD Dr. Ilian Jelezarov

Zürich 2007

Summary

The Max protein (Myc associated factor X) belongs to the Myc/Mad/Max network of basic helix-loop-helix-transcription factors controlling cell metabolism, proliferation and differentiation. The network includes, next to Max, members of the Myc and Mad families, the Mad-related protein Mnt, and Mga. Max appears to have a central role in concerting the action within the network, since it is the obligate heterodimerization partner of Myc, Mad and Mnt, thereby promoting recognition of the enhancer box (E-box) sequences in the promoters of a large set of genes, and triggering various activities of the network. Myc/Max heterodimers activate genes involved in cell growth and proliferation. Myc deregulation and overexpression is associated with apoptosis, genetic instability and malignant transformations. The Mad/Max heterodimer antagonizes most of Myc/Max transcriptional activities by recruiting corepressors with HDAC activity. The Mnt/Max heterodimer also has a tumour suppressor activity. While Myc, Mad and Mnt proteins have short half-lives and their expression levels are strictly controlled by diverse signal transduction cascades, Max itself has a significantly longer half-life and is stably and ubiquitously expressed. Being transcriptionally inert and lacking the domains necessary for recruiting coactivator or coreressor protein complexes, Max is thought to modulate cell growth by competing for E-box with Myc/Max, Mad/Max or Mnt/Max. It is still unknown whether the competition proceeds at the level of heterodimer-*versus*-homodimer E-box binding affinity discrimination, or else $(Max)_2$ binding to DNA reduces the concentration of Max monomers available for heterodimerization with other family members.

The presented thesis is devoted to characterization of the biophysical properties of the full length gene product of the Max p21 isoform and the energetics of E-box recognition by Max p21. Specific E-box binding by Max is mediated by conserved side chains from the basic region. Structural studies have identified a conserved and quasi-symmetric recognition pattern within the 5'-CACGTG-3' target, including four specific hydrogen bonds to DNA bases and a number of non-specific electrostatic contacts with the phosphate backbone. In addition, residues from the loop region closely approach the duplex and are seemingly involved in binding. So far, only the binding affinity of short constructs encompassing the b-HLH-LZ core domain has been determined. In the present thesis I use mainly calorimetry to characterize thermodynamically the site-specific binding of the complete Max p21 isoform to a 21 base pair DNA duplex containing the E-box sequence. The work provides for the first time a reliable estimate of the dissociation constant at the physiologically relevant temperature. Max p21 binds to the DNA target site with low nanomolar affinity at 37 °C. The association is driven by a large exothermic effect, which is partly compensated by entropic factors. The energetic contribution of seven, highly-conserved residues (six in the basic region and one in the loop) that contact the DNA to binding affinity was probed by alanine scanning mutagenesis. Removal of each of the contacts made by His 18, Glu 22, Arg 23, Arg 25 and Arg 50 reduces the E-box binding affinity by a factor of 15 to 160. The Arg 26 side chain confers very substantial stabilization to the Max p21-E-box complex, while Asn 19 appears energetically unimportant. Partitioning of the free energy of binding in terms of enthalpy and entropy reveals the complicated, context-dependent thermodynamic signature of particular protein-DNA contacts. Altogether, the mutational analysis points to the central role which the persistent helical turn spanning the C-terminus of the basic region and the start of helix H1 plays to tight DNA binding. An increase in DNA-binding affinity upon stabilization of the leucine zipper domain is demonstrated. Structural rearrangement within the adjacent HLH possibly leads to formation of additional protein-phosphate backbone contacts. The results provide further thermodynamic support to the fact that the N-terminal and C-terminal protein segments outside the core b-HLH-LZ domain appear to be largely unstructured and not participating in DNA binding. Finally, experiments with a strongly destabilized Max p21 variant appear to contradict previous results supporting high-affinity E-box recognition by Max monomers.

Zusammenfassung

Das Max (Myc assciated factor X) Protein gehört zum Myc/Max/Mad Transkriptionsfaktor-Netzwerk, welches den Metabolismus, die Proliferation und die Differenzierung einer Zelle kontrolliert. Das Netzwerk umfasst, nebst Max auch die Proteine der Myc- und der Mad-Familien, das Mad verwandte Protein Mnt und das Mga. Max scheint eine zentrale Rolle in diesem Netzwerk zu spielen, denn es ist der obligate Heterodimerisierungspartner der Myc, Mad und Mnt Proteine und ermöglicht ihnen so erst die Erkennung der enhancer box (E-box)-Sequenzen in den Promotoren einer Vielzahl wichtiger Gene, welche das Zellwachstum und die Zellteilung steuern. Myc als Heterodimer mit Max aktiviert die Gene, die für Zellwachstum und –Teilung eine grosse Bedeutung haben. Myc Deregulation und Überexpression führt zur Apoptose, zu genetischer Instabilität und in vielen Fällen zu maligner Transformation. Das Mad/Max Heterodimer steht dem Myc/Max diametral gegenüber: Es antagonisiert seine Funktion durch Repression der HDAC Aktivität. Das Mnt/Max Heterodimer, das wurde schon in einigen Studien gezeigt, besitzt hochpotente Tumorsuppressor-Aktivität. Die Myc, Mad und Mnt Proteine haben kurze Halbwärtszeiten und ihre Expressionsmenge wird streng von verschiedenen Singaltransduktionskaskaden kontrolliert. Max selbst ist relativ stabil in der Zelle und wird überall und in konstanter Mengen exprimiert. Max ist transkriptionell inert, die Domänen, die für die Rekrutierung der Co-Aktivator und Co-Repressor Proteinkomplexe benötigt werden, fehlen. Die Genrepression durch Max geschieht höchstwahrscheinlich durch Kompetition von Myc, Mad und Max um die Hetero- beziehungsweise Homodimerisation oder um die Bindung an die DNA. Jedoch ist die Ebene der Kompetition noch unbekannt.

Die folgende Arbeit ist der biophysikalischen Charakterisierung des kompletten Genproduktes Max p21 und der Energetik der Max p21-Ebox Bindung gewidmet. Die spezifische Bindung an die E-box durch Max wird durch konservierte Aminosäurereste in der basischen Region vermittelt. Gemäss der Kristallstruktur besitzt der Max-E-box Komplex ein quasi-symmetrisches, konserviertes Erkennungsmuster im 5'-CACGTG-3' Erkennungsmotiv, inklusive vier spezifische Wasserstoffbrückenbindungen und eine Vielzahl unspezifischer elektostatischer Kontakte mit den Phosphaten der DNA. Zusätzlich kommen Aminosäuren in der „loop"-Region der DNA sehr nahe und scheinen ebenfalls eine bedeutende Rolle in der Protein-DNA Interaktion zu spielen. Bis heute sind nur Bindungskonstanten eines verkürzten Max Homodimers, das nur die bHLHLZ Region beinhaltet, gemessen worden. In der folgenden Arbeit charakterisiere ich hauptsächlich mit kalorimetischen Methoden die Protein-DNA Interaktionen des natürlichen Max p21-Isoform in seiner vollen Länge mit einem 21 Basenpaar langem DNA-Stück, welches die E-box Erkennungssequenz beinhaltet. Diese Dissertation liefert zum ersten Mal eine verlässliche Messung der Dissoziationskonstante vom Max p21-DNA Komplex unter physiologisch relevanten Bedingungen. Bei 37°C bindet Max p21 mit hoher Affinität im Bereich 10^9 an die E-box. Die Assoziation weist einen wesentlich hohen exothermen Effekt auf, der jedoch teilweise von entropischen Einbussen kompensiert wird. Der energetische Effekt der sieben hoch-konservierten Aminosäuren (sechs in der basischen Region, einer im loop), welche die DNA kontaktieren wurde durch jeweilige Alanin-Mutationen untersucht. Das Entfernen der einzelnen DNA-Kontakte durch His 18, Glu 22, Arg 23, Arg 25 und Arg 50 vermindert die E-box-Bindungsaffinität um einen Faktor zwischen 15 und 160. Arg 26 ist für die Stabilisierung des Max p21-Ebox-Komplexes von grosser Bedeutung. Im Gegenteil dazu Asn 19, das energetisch kaum eine Rolle einzunehmen scheint. Teilen wir die freie Gibbs-Energie der Bindung in entropische- und enthalpische Faktoren auf, stellen wir eine komplizierte und vor allem kontext-abhängige thermodynamische Signatur der einzelnen E-box-kontaktierenden Aminosäuren fest. Alles in allem hat die Mutationsanalyse gezeigt, dass die Helizität des C-Terminus in der basischen Region für eine starke Protein-DNA-Assoziation sehr wichtig ist. Die stabilisierung des Max-E-box-Komplexes aufgrund eines stabileren Leuzinzippers wird in der folgenden Arbeit ebenfalls gezeigt. Konformationsänderungen in der HLH Region sind vermutlich dafür verantwortlich, dass bei stabilerem Leuzinzipper mehr elektostatische Protein-DNA-Kontakte geknüpft werden. Die folgende Arbeit bietet ausserdem thermodynamische Unterstützung der These, welche besagt, dass die N- und C-terminalen Regionen ausserhalb des bHLHLZ unstrukturiert sind und in der Protein-DNA-Interaktion unbeteiligt. Der Schlussteil dieser Arbeit beinhaltet Experimente mit einem stark destabilisierten Max p21 Homodimer. Die Resultate widersprechen der These, dass Max Monomere mit hoher Affinität an die E-box DNA binden.

1 Introduction

1.1 Protein-DNA interactions: general remarks

Knowledge about the mechanism of DNA recognition by proteins provides important insights of the way living cells function. DNA carries the genetic information for all vital biological processes, but the realization of this information comes as the consequence of the action of the protein molecular machinery. Proteins bind to DNA to fulfill diverse tasks: transcription, chromatin packing and re-shaping, replication, restriction, DNA repair, etc. Structural and biophysical studies of DNA-protein interactions are of fundamental importance to understand the molecular basis of biological function. Moreover, accumulated information about the structural, energetic and dynamic aspects of protein-DNA interactions may guide the development of strategies and practical approaches aimed at biomedical and biotechnological applications. In spite of the ever growing number of structural and biophysical studies of DNA-protein complexes, no universal structural or energetic code of DNA-protein recognition has yet emerged. Only some loose and very general relationships have been found. The majority of DNA binding proteins have positively charged binding surfaces providing, in principle, favorable electrostatic interactions with the negatively charged DNA. However, electrostatic interactions are not the most important energetic contribution to affinity and specificity. Being largely non-specific in nature electrostatics only seems to help speeding up association trough electrostatic steering. The main driving forces for protein-DNA binding are van der Waals interactions and hydrogen bonds. Indeed, known structures of protein-DNA complexes exhibit very good steric complementarily of non-polar surfaces and favorable orientation of hydrogen bond donors and acceptors. The specificity and strength of protein-DNA complexes is often achieved by conformational adaptation transitions of protein domains, DNA bending, or both. Water molecules which are trapped at the protein-DNA interface frequently optimize the steric complementarity. A research program combining structural, kinetic and thermodynamic characterization of DNA protein interactions will deepen our knowledge about the fundamental principles of macromolecular recognition, and will also facilitate progress in pharmaceutical and gene therapeutic approaches toward modulation of the function of protein-DNA complexes. Binding could be abolished, strengthened, specificities could be changed or new specificities could be designed. In this context, understanding the structural basis and determinants providing affinity and specificity of DNA-protein complexes is of special biological, medial and industrial interest.

1.2 Thermodynamics of macromolecular recognition with emphasis on protein-DNA binding

Finding links between sequence, structure and function is a long lasting endeavor of contemporary biochemistry and biophysics. Energetics and dynamics become more and more recognized as key pieces of information required for a comprehensive description of biological processes. In spite of the progress that has been made in developing a theoretical framework for computational analysis of biomacromolecular interactions, it is not possible to make quantitative conclusions on the stability of a particular protein-DNA complex from structural data alone. Moreover, as the results described in this thesis demonstrate, it is not possible to draw definitive conclusions how structural changes upon mutations are reflected in changes in the thermodynamic signature of binding. Thermodynamics describes binding reactions in terms of temperature, pressure, free energy, enthalpy and entropy, thus providing information of whether and why these reactions are going to occur. Thermodynamics has proved its power in understanding the properties and behavior of simpler systems like gases, liquids and small molecule reactions, and develops more and more as a useful tool in the research of more complicated systems like protein-DNA complexes. Protein-DNA association is energetically driven by contributions from non-covalent bonds: van der Waals contacts, oriented hydrogen bonds, salt bridges, and long range electrostatic interactions, the strength of all of them depending on the environmental variables (temperature, pH, osmotic pressure and salt and other additives concentration). Next to surface complementarity and geometrical restraints, the energetic content of a particular non-covalent bond is modulated by the degree of dehydration of the constituent groups at the protein-DNA interface. Conformational adaptation of protein and DNA is linked to energetic expenditures. The surface of protein and DNA is hydrated and ions might be tightly bound to solvent-exposed groups. Water and other solutes are sometimes trapped at or expulsed from the complex interface and make contributions to the strength of binding. Proteins and the DNA are flexible molecules and can experience significant thermal fluctuations. All mentioned forces and energetic effects are finely tuned. General insights into the energetic signature of binding reactions come from knowledge of the enthalpy, entropy, free energy and heat capacity changes accompanying binding, and the dependence of these parameters on the temperature and other physical and chemical variables. Furthermore, the changes of thermodynamic parameters upon mutation might reveal the energetic importance of discrete protein-DNA contacts. Combination of the thermodynamic description with structural analysis provides invaluable information on the determinants of binding affinity and specificity. In the following I briefly summarize the formalism relevant to thermodynamic description of binding reactions and some general observations concerning the sign and magnitude of thermodynamic parameters pertinent to protein-DNA association.

1.2.1 The Gibbs free energy

The free Gibbs energy (ΔG) is the thermodynamic parameter which provides most fundamental information about thermodynamic systems. It describes the overall change in free energy of any chemical reaction and is the direct indicator of the direction of spontaneous chemical transformation. Chemical reactions proceed spontaneously in direction to the state which has the lowest Gibbs free energy. Furthermore, the Gibbs free energy indicates the equilibrium population of molecular species at certain fixed conditions. There is no basic distinction between a chemical reaction and association between macromolecules. If ΔG is known, the population of free macromolecules and their complex is rigorously defined. Hence, the magnitude of ΔG is a direct measure for the stability of a non-covalent complex. ΔG can be experimentally determined from measurement of the equilibrium binding constant, K_{eq}, using the relationship:

$$\Delta G = -RT \ln K_{eq} \tag{1.1}$$

K_{eq} itself is defined as:

$$K_{eq} = \frac{[PD]}{[D][P]} \tag{1.2}$$

where the square brackets indicate the equilibrium concentrations of protein (P), DNA (D) and protein-DNA complex (PD). ΔG is the sum of many entropic and enthalpic terms, which comprise the overall enthalpy change (ΔH) and the overall entropy change (ΔS) according to the Gibbs-Helmholtz-equation:

$$\Delta G = \Delta H - T\Delta S \tag{1.3}$$

Both ΔH and ΔS depend on the temperature. If the heat capacity change (ΔC_p) of the considered process is non-zero, at constant pressure:

$$\Delta H(T) = \Delta H(T_R) + \int_{T_R}^{T} \Delta C_p dT \tag{1.4a}$$

$$\Delta S(T) = \Delta S(T_R) + \int_{T_R}^{T} \frac{\Delta C_p}{T} dT \tag{1.4b}$$

Therefore ΔG is also a function of the temperature:

$$\Delta G(T) = \Delta H(T_R) - T\Delta S(T_R) + \Delta C_p \left[T - T_R - T \ln\left(\frac{T}{T_R}\right)\right] \tag{1.5}$$

If ΔC_p is itself temperature-independent, the integrated form of the Gibbs-Helmholtz equation is:

$$\Delta G(T) = \Delta H(T_R) - T\Delta S(T_R) + \Delta C_p \left[T - T_R - T \ln\left(\frac{T}{T_R}\right)\right] = \Delta G(T_R) + \Delta C_p \left[T - T_R - T \ln\left(\frac{T}{T_R}\right)\right] \tag{1.6}$$

In the above equations T indicates any arbitrary temperature, and T_R is an appropriate reference temperature, at which the enthalpy change $\Delta H(T_R)$ and the entropy change $\Delta S(T_R)$ are known. Equation 1.6 is generally valid and can be used to calculate the stability of macromolecules and of protein-ligand complexes over a certain temperature range. In practical terms, the temperature dependence of the affinity of protein-ligand complexes is determined from direct measurements of K_{eq} at different temperatures. Alternatively, $T\Delta S$, ΔH, and ΔC_P at a given temperature (T_R) must be known, in order to use equation 1.6. It is important to note that the formalism outlined above is not an end it itself. In numerous cases, the binding affinity can not be measured at the physiologically relevant temperature (mostly 37°C). Therefore, ΔG must be extrapolated from the experimentally accessible temperature interval according to Equation 1.6. In terms of K_{eq} the extrapolation is done by:

$$K_{eq}(T) = K_{eq}(T_{exp}) \exp\left\{ -\frac{\Delta H_{exp}}{R}\left(\frac{1}{T_{exp}} - \frac{1}{T}\right) + \frac{\Delta C_{p,exp}}{R}\left(\ln\frac{T}{T_{exp}} + \frac{T_{exp}}{T} - 1\right)\right\} \tag{1.7}$$

R is the universal gas constant (8.314 kJ $mol^{-1}K^{-1}$); T_{exp} is the experimentally accessible temperature (usually below 30°C). In macromolecular interactions the heat capacity changes upon complex formation are significant, so that enthalpic and entropic contributions have pronounced temperature dependence (*1, 2*). Since their signs are often opposite, ΔH and ΔS effectively compensate each other, making the temperature dependence of ΔG small. This phenomenon is called entropy-enthalpy-compensation. It can be shown that $\Delta H \cong T\Delta S + \Delta C_p(T_S - T_H)$, where T_S and T_H are the temperatures where ΔS and ΔH equal zero, respectively (*3*). There is, indeed, a linear correlation between ΔS and ΔH, since the second addend of the right–hand side of the above expression is a constant for each particular protein-DNA system. It has been speculated that the strong enthalpy-entropy-compensation observed in protein folding and macromolecular binding is grounded in the specific chemical features of the macromolecular surface and the important role water rearrangement plays in macromolecular binding (*4-6*). However, the physical origin of this phenomenon is not fully understood.

A statistical survey of protein-DNA interactions in the proNIT database provides an average ΔG value of 20-60 kJ mol^{-1} for protein-DNA associations in the range of 0°C-40°C and pH 6-8. Hence, equilibrium constants K_{eq} at 25°C vary from 10^4 to 10^{10} M^{-1}.

1.2.2 The enthalpy change

The enthalpy change provides information about the overall energy of non-covalent bonds newly formed between protein and DNA. At constant pressure it can be written as:

$$\Delta H = -R\left[\frac{\partial \ln K_{eq}}{\partial \ln \frac{1}{T}}\right] \qquad (1.8a)$$

or alternatively in an equally popular notation:

$$\Delta H = RT^2 \frac{d \ln K_{eq}}{dT} \qquad (1.8b)$$

It is clear that the enthalpy change of binding can be obtained by the temperature dependence of the binding constant at constant pressure. However, since the temperature variation of K_{eq} is typically not very strong, and the measurements contain sizeable experimental error, ΔH estimates obtained from temperature derivatives of K_{eq} (equations 1.8a and 1.8b) are not very precise. Much more reliable ΔH values are available from isothermal calorimetry experiments, where the heat effect of the binding reaction is measured directly.

The overall enthalpy change of a binding reaction represents the sum of the enthalpy of non-covalent protein-DNA bonds, the enthalpy of any conformational changes accompanying binding, the enthalpy of protonation/deprotonation reactions, and the enthalpy of formation/breakage of water-macromolecule contacts. It is very important to note that the enthalpy of a binding reaction is evaluated in respect to the reference state, which is usually the dissociated state. Considering the enthalpic contribution of van der Waals contacts and hydrogen bonds, this means that the favorable enthalpy of any bond *per se* is offset by the unfavorable enthalpy of dehydration of the participating groups, when they are transferred into the tightly packed protein-DNA interface. In many protein-DNA complexes water molecules and clusters of water molecules bridge protein and DNA (*2, 7*). Extreme cases are known, where almost all protein-DNA contacts are mediated by water (*8, 9*). Water-protein and water-DNA hydrogen bonds are thought to have a significant enthalpic impact. Furthermore, the overall enthalpy change is modulated by the pronounced enthalpic effect of conformational adaptation transitions. In some cases, entire DNA-binding domains are unfolded in the absence in DNA but obtain regular and stable conformation when the protein-DNA complex is formed (*10*). Such folding transitions are linked to heat release, making the overall enthalpy of the binding reaction more negative. A statistical survey of protein-DNA binding thermodynamics data collected in the proNIT data base for 80 protein-DNA complexes indicates an average value of $\Delta H = -21$ kJ mol^{-1} at 25°C and pH 6-8. Interestingly, there is no correlation between the sign and magnitude of ΔH and the strength of binding: endothermic binding (unfavorable enthalpy change) is often quite strong. In view of the above said, however, it is not a surprise that the range of ΔH variation is extremely wide. For example, binding of λcI repressor to its target duplex is enthalpically favorable with $\Delta H = -100$ kJ mol^{-1} while binding of the TATA-box binding protein to TATA-box DNA is enthalpically disfavored by as much as 125 kJ mol^{-1} at the same temperature (*11*). The cited extreme cases merely illustrate that the balance of enthalpic effects at protein-DNA interfaces similar size and chemical composition could be extremely different.

1.2.3 The entropy change

Formally the entropy change of a binding reaction is calculated from non-calorimetric data by the van t' Hoff expression:

$$\Delta S = -\frac{d\Delta G}{dT} \tag{1.9a}$$

or from calorimetric data as:

$$\Delta S = \frac{\Delta H - \Delta G}{T} \tag{1.9b}$$

As discussed above, protein-DNA complexes can be very stable even if the enthalpy change opposes binding overall (endothermic binding). Formation of such complexes, therefore, is driven entirely by favorable entropy changes, at least in some temperature range. At first glance this is surprising, since the entropy is intuitively regarded as measure of the degrees of freedom, which are expected to decrease upon formation of a bimolecular complex. However, for the entropic balance of the reaction one should consider all processes having entropic consequences. The overall entropy, ΔS_{tot}, can indeed be expressed as a sum of different contributions:

$$\Delta S_{tot} = \Delta S_{rt} + \Delta S_{conf} + \Delta S_{hyd} + \Delta S_{cat} \tag{1.10}$$

The term ΔS_{rt} represents the change in rotational and translational degrees of freedom. It is negative and opposes binding at all temperatures. However, newer experimental work suggests that its magnitude is close to the cratic entropy loss, which is a small number in comparison to all other energetic contributions (*12*). The term ΔS_{conf} includes all changes of conformational entropy within the protein-DNA complex with reference to the dissociated state. ΔS_{conf} is typically also negative, even if binding proceeds without conformational adaptation, since protein side chains and backbone become “frozen” at the binding interface. The magnitude of ΔS_{conf} is strongly dependent on the exact degree of structural rearrangements taking place. For example, cases are known where some parts of a DNA-binding protein domain become in fact more disordered in the complex than in the free protein, and this entropic gain (at least partly) compensates the entropic loss typically occurring at the protein-DNA interface (*13*). The last two terms of equation 1.10 are in fact linked to entropy gain, arising from the favorable entropy of dehydration of a tightly packed intermolecular interface (ΔS_{hyd}) and from the likewise favorable expulsion of cations, which are bound to DNA backbone phosphates, but are replaced by positively charged protein groups in the complex (ΔS_{cat}). The sum of ΔS_{hyd}

and ΔS_{cat} represents a significant driving force of the binding reaction, even if trapping of water molecules at the protein-DNA interface is entropically unfavorable. It should be noted that ΔS_{rt} and ΔS_{conf} are temperature independent (in a first approximation) whereas ΔS_{hyd} is strongly temperature dependent. It follows that depending on the balance between the terms of equation 1.10 and the temperature the total entropy change might drive binding. Taking again as the example the TATA-box binding protein and λcI, the unfavorable binding ΔH (+120 kJ mol^{-1}) of TATA-box recognition is overcompensated by ΔS = +550 J mol^{-1} K^{-1} (TΔS ~ 165 kJ mol^{-1} at 25 °C) while the large exothermic effect of λcI binding to its operator DNA (–100 kJ mol^{-1}) is opposed by a decrease in entropy of –180 J mol^{-1} K^{-1} (TΔS ~ –54 kJ mol^{-1} at 25 °C). It is clear that both complexes are equally stable at room temperature (ΔG = –44 kJ mol^{-1} for TATA-box binding protein-DNA complex and ΔG = –46 kJ mol^{-1} for the λcI-DNA complex) but for entirely different thermodynamic reasons. For the most protein-DNA association reactions ΔS in the range –200 to +200 J mol^{-1} K^{-1} has been measured at 25 °C and neutral pH.

1.2.4 The heat capacity change

The heat capacity change is a measure of the temperature dependence of the enthalpy and entropy changes at constant pressure.

$$\Delta C_p = C_p{}^{comp} - (C_p{}^{pr} + C_p{}^{DNA}) = [\partial(\Delta H)/\partial T] = [T\partial(\Delta S)/\partial T] \qquad (1.11)$$

where $C_p{}^{pr}$, $C_p{}^{DNA}$ and $C_p{}^{comp}$ are the heat capacities of the protein, DNA and their complex respectively. Experimentally, ΔC_p is most reliably calculated from calorimetric data by the Kirchhoff equation:

$$\Delta C_p = \frac{d\Delta H}{dT} \qquad (1.12)$$

In principle the heat capacity change is accessible from non-calorimetric data as well:

$$\Delta C_p = \frac{d^2\Delta G}{dT^2} \qquad (1.13)$$

However, the error is significant. The heat capacity change has attracted special attention because its accurate determination is critical for a rigorous description of the enthalpy, entropy and free energy changes of a binding reaction. The experimentally obtained values of the heat capacity changes of protein-DNA interaction are negative when using the free components as the reference state (Equation 1.11). Similarly to the other thermodynamic parameters (sections 1.2.2 and 1.2.3), ΔC_p can be partitioned into contributions arising from intermolecular interactions in vacuum, conformational changes and hydration effects. The large negative ΔC_p of binding is ascribed mostly to the loss of bound water molecules from the interacting surface areas in the complex (*14*). Small-scale conformational changes that the protein and DNA might undergo, as well as intermolecular interactions (van der Waals, electrostatic), are typically considered as smaller contributors. Their qualitative (positive or negative) and quantitative effects on the heat capacity change are difficult to predict. The magnitude of ΔC_p should therefore roughly correlate with the energetic cost of dehydration of molecular surface (*12, 15*). The heat capacity of dehydration itself is a function of temperature, but this temperature dependence is small and can be neglected in the temperature interval of typical binding experiments. In a wide temperature range the contributions from dehydration of polar groups is positive: dehydration of aromatic and aliphatic groups is linked to ΔC_p decrease. The heat capacity change (and also the enthalpy and entropy changes) can be estimated by semi-empirical methods from the structure of the complex and its components in isolation (*12, 15*), although these estimates may be ambiguous (*4, 16*). These correlations, however, are only valid if the interaction species are considered as rigid bodies and their surfaces are geometrically complement, as to exclude all hydration water from the interface. Values of ΔC_p for protein-DNA interaction tend to be larger than predicted from the burial of polar and non-polar surface alone (*17-21*). The “excess” ΔC_p arises from large-scale conformational transitions (including also DNA bending) (*17*), entrapment of water molecules at the binding interface (*13*) or substantial, non-parallel changes of the heat capacities of the associated or dissociated state of the system (*21*).

1.3 The Max protein as a model to study the energetics of site-specific DNA recognition by dimeric transcription factors

1.3.1 Biology of transcription factors

Recently, the entire genome of a number of organisms has been sequenced (*22, 23*). However important, the genome sequence itself could not describe any organism as a functional and unique entity (*22, 23*). For example, comparing the genomes of different organisms revealed surprising similarity of genetic content (*23, 24*). The genome of vertebrates consists of only twice as many genes as that of invertebrates, the increase being mostly due to gene duplication events (*23*). Morphological and behavioral complexity is not only a function of the number of genes but also of the regulation of the genes, mostly through regulation of gene transcription. It is now known that humans transcriptional machinery consists of approximately 3000 transcription factors regulating about 26000 genes (*23*). Understanding the regulation of genes, especially of those involved in diseases, will help to understand the disease itself and builds the basics to find an appropriate cure.

It is out of the scope of the present work to discuss in detail the complicated mechanisms leading to tight regulation of gene transcription. As a mere illustration of the complexities, I briefly sketch the current picture about the biological impact of c-Myc transcription factor. Myc is up-regulated and over-expressed in many different types of cancers (*25*), most of which are very aggressive and lead quickly to death. In Neuroblastoma, a kidney tumour, often developing during childhood in humans, over-expression of a Myc subtype, N-Myc, stimulates cell growth continuously. There are also Myc-independent forms of Neuroblastoma. The prognosis is highly dependent on whether the cancer is due to Myc up-regulation or not. In the Myc-dependent case the average 5-year- survival rate of the patient is only 5%, while in the Myc independent case there is good chance for survival, if the cancer is diagnosed early. Essentially the same picture is encountered in small cell lung cancer, where the isoform L-Myc is up-regulated. Myc up-regulation (and cancer development for that matter) is achieved via different routes. Frequently de-regulation of signal transduction cascades are the cause of up-regulation of transcription factors, as exemplified by the Bcr-Abl protein being responsible of Myc over-expression (*26, 27*).

1.3.2 The Myc-Max-Mad transcription factor network

The Myc-Max-Mad network comprises a group of conserved protein families of transcription factors, whose distinct interactions result in gene-specific transcriptional activation or repression. Diverse biological effects are brought about by the inherent property of the network members to form different dimeric species (*28-30*). The Max protein (Myc associated factor X) has a central role in concerting the actions within the network, since it is capable to homodimerize, as well as to heterodimerize with other proteins (*31*). In contrast, Myc, Mad, Mnt and Mga proteins are extremely unstable as homodimers and need Max as the obligate binding partner (*29, 30, 32, 33*). Homo- and heterodimers compete for binding to the same enhancer box (E-box) sequence of a large set of genes. The targeted E-boxes are either high-affinity (CACGTG) or low-affinity (CACATG) and could be separated by as much as 2 kb (*34*).

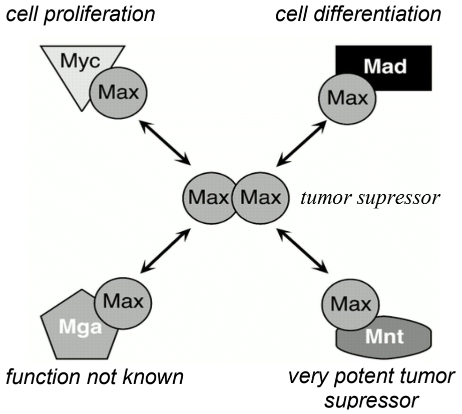

Figure 1.1. Schematic representation of the interactions within the Myc-Max-Mad network. Max is the only family member capable to form homodimers. Myc, Mad, Mga and Mnt proteins form heterodimers with Max. The biological effect of different dimeric species is indicated.

Identification of the Myc target genes has turned to be a difficult task. Nevertheless, recent work has led to identification of a large set of genes, which are regulated by Myc-Max, Mad-Max, Mnt-Max, Mga-Max, or Max homodimer itself. Most of the high affinity binding sites for the transcription factors are clustered together in the human genome. The Myc-Max-Mad targets belong to several diverse functional categories Next to regulating cell division, proliferation and differentiation, thus regulating the cell cycle (*29*), the network action has been implicated in vital processes like signal transduction, regulation of translation, chaperone expression, etc (*29, 30, 32, 33*). Figure 1.1 shows a schematic overview of the interactions within the Myc-Mad-Max network. A list of well characterized gene targets includes cyclin D, Cdc25a and CDK4 (cell cycle activation), Hsp70 and Hsp90 (chaperones), E2F1 and EGR3 (signal transduction), EIF4E, DDX18 and LDH-A (cell growth), and Mucin 1 (ribosomal protein).

In the following sections, a brief review of the biological effects of members of the Myc-Mad-Max network is given. The biology, structure and biophysical properties of the Max-Max homodimer, the object of the present study, is presented in more detail in Section 1.3

Myc-Max

The Myc Protein (Myelocytoma) was discovered in the 1980ties in simian myeloid leukemia (*35*). Amplification of the Myc-family genes (c-Myc, L-Myc, N-Myc) is one of the most frequent causes of human cancer. Miss-regulation of cell signaling is not the only cause of Myc over-expression. It has been shown that chromosome translocation leads to disruption of a stringent cis-acting control element of the myc gene.

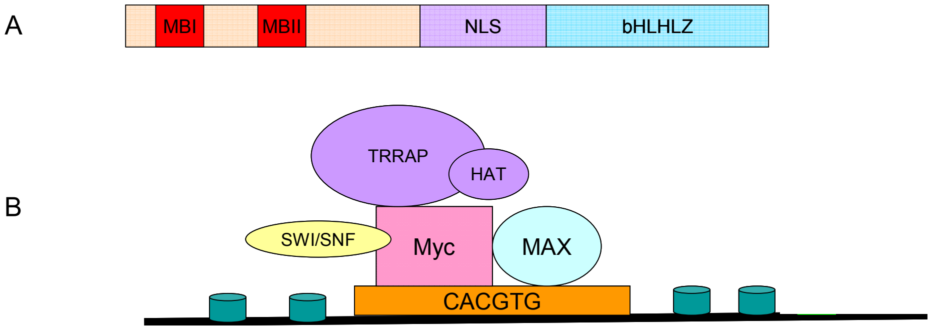

Figure 1.2 A**,** Sequential domain organization of the c-Myc protein. The transactivation domain containing the highly conserved Myc Box I and Myc Box II (MBI, MBII shown in red) is located N-terminally, followed by the nuclear localisation signal (NLS, violet) and at the basic helix-loop-helix/leucine zipper domain, which is responsible for dimerization and site-specific DNA binding (bHLHLZ, cyan). B, Activation of the genes responsible for cell growth by Myc-Max. This process requires the transactivation domain which recruits the Histone deacetylase complex (TRRAP in complex with HAT, violet) and the auxiliary factor SWI/SNF to relax the chromatin structure and open up the chromatin. Promotors are shown as green cylinders.

Normally this control element keeps a low level of Myc in the cell. Recent estimations suggest that there are more than 2500 binding sites in the human gene for Myc-Max transcription factor stimulating anabolic cellular processes (*34*). Myc-Max is involved in the regulation of RNA polymerases I and III. Myc over-expression sensitizes the cell to apoptosis by direct regulation of pro-apoptotic genes (*36*). However, Myc-induced tumours do not undergo apoptosis. It is believed that some Myc induced tumour cells have lost their apoptotic response to Myc or are not programmed to respond in this way

(*37*) The domain organization of the Myc gene product and the mechanism of gene activation by Myc-Max are presented in Figure 1.2.

Mad-Max

The Mad family (Mitotic absence deficient) comprises Mad1, Mxi1, Mad3 and Mad4 proteins (*29, 38, 39*). All Mad proteins heterodimerize with Max, thus repressing cell growth and stimulating entry into the G0-phase and differentiation. Over-expression of Mad interferes with cell growth and proliferation.

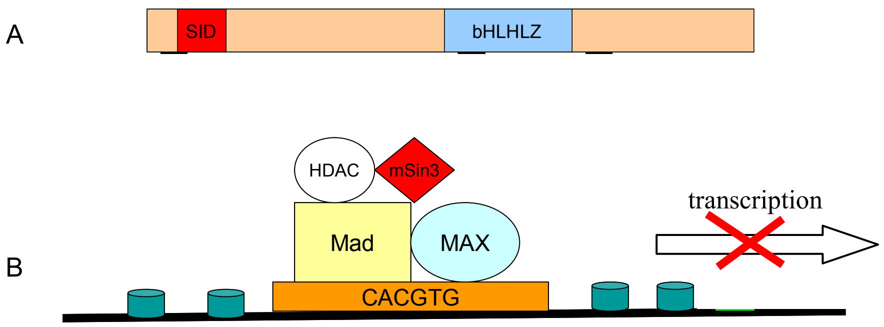

Figure 1.3. A, Sequencial domain organization of Mad1. The N-terminal Sin3 interacting domain (SID), comprising residues 8-20 within the transactivation domain is shown in red. The DNA-binding and dimerization bHLHLZ domain is in blue. B, Transcriptional repression by the Mad-Max heterodimer. The transactivation domain recruits the repressor histone acetylase complex (HDAC and mSin3; white and red shapes, respectively) to repress the corresponding genes. Green cylinders represent promotors

Expression of Mad blocks malignant transformation. Furthermore, Mxi1 is able to block the growth of Glioblasoma multiforme. Hence, Mad proteins antagonize Myc. Interestingly; Mxi1 has been mapped to chromosomal regions which are often deleted in tumours. The domain organization of the Mad1 gene product and the mechanism of gene activation by Mad-Max are presented in Figure 1.3.

Mnt-Max

The Mnt protein (Melanostic neurektodermal tumours), known also as Mad6, Rox or Mxd6, probably serves as a global general antagonist of Myc-Max. Mnt knockout mice showed severe symptoms of Myc over-expression, and died 24 hours after birth (*40, 41*). Both, Mad and Mnt protein families are Myc antagonists (*40, 41*). However, Mnt is a much more potent tumour suppressor than Mad. While Mad- deficient mice are viable, fertile and not tumour prone (*40, 41*), Mnt-deficient mice show severe tumour genesis and Mnt deficiency is lethal. The exact reason for the observed differences in tumour suppression potency of Mnt and Mad are still unclear.

Mga-Max

Hurlin et al., identified a novel Max-interacting protein named Mga (*42*). Like Myc, Mad and Mnt, Mga heterodimerizes with Max and recognizes E-box sequences. In contrast to other known members of the network, Mga contains a second highly conserved DNA binding motif, the T-box domain, which is characteristic for the Tbx family of transcription factors. The presence of both, a T-domain and a E-box-targeting b-HLH-LZ domain, suggests (i) that the transcriptional and biological activities of Mga are probably more complex than those of other members of the b-HLH-LZ and T-domain families, and (ii) that Mga might represent a target point for concerted regulation of the activities of the Myc-Mad-Max network and the T-protein network.(*42*). From the still limited information available to date, it has been inferred that Mga may impinge on the function of other Max-interacting proteins. For example, transcription assays revealed a suppression of Myc-dependent transformation by Mga in a dose-dependent manner (*42*). Interestingly Mga suppression of Myc activation was less potent than that caused by Mad or Mnt.

1.3.3 Biology, structure and biophysics of the Max protein

As already pointed out, Max appears to be the key player in the coordinated action of the Myc-Mad-Max network. In contrast to other network members, the protein is stably and ubiquitously expressed at relatively high level, and has a long half-life (*29*). The Max homodimer represses the transcription of Myc-up-regulated genes that are involved in cell growth and proliferation, and is therefore the principal functional antagonist of Myc.

1.3.3.1 Isoforms

In human cells Max is expressed in two common isoforms: Max p22 (160 amino acids) and Max p21 (151 amino acids). The two differ in a 9 amino acid long insert in the N-terminal segment of Max p22 (see Figure 1.4).

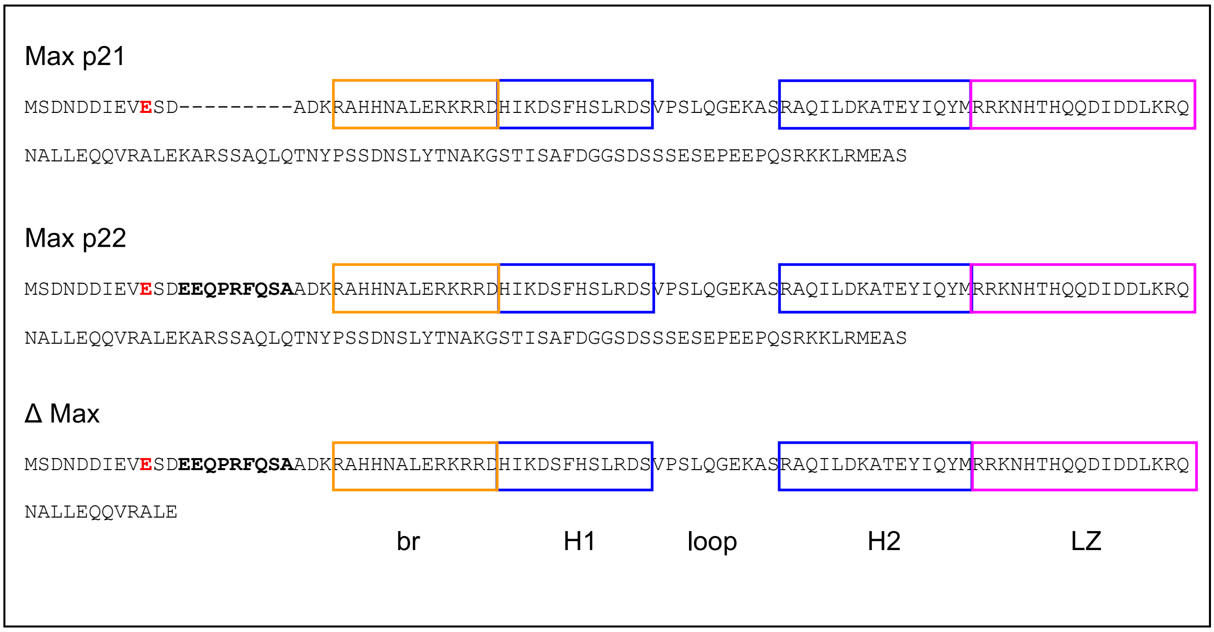

Figure 1.4. The isoforms of Max. Isoforms p22 and Δ Max contain an N-terminal 9 amino acid long insert (in bold). The sequences encompassing the secondary structure elements forming the helix-loop-helix/leucine zipper domain (HLHLZ) are boxed. Blue boxes, α-helices H1 and H2; magenta boxes, leucine zipper (LZ). The loop region connecting helix H1 and helix H2 is also indicated. The basic region (br) delineated with broun boxes. The cleavage site for caspase 5 (Glu 9) is shown in red.

A third, unusual isoform named ΔMax (103 amino acids) has a truncated C-terminus, thus lacking the nuclear localization signal. There is some evidence that ΔMax plays a role in Myc-Ras co-transformation (*43*). Considering the main isoforms p21 and p22, the role of the insert in their presumably different biological effect(s) is yet not known. It is firmly established that Max p22 is cleaved at Glu 9 (Max p22 sequence) by caspase 5 during Fas-mediated apoptosis (*44*). Max p21 is also cleaved but less efficiently than the longer isoform. It was suggested that the reason most probably lies in structural differences in the N-terminal segment, which is devoid of stable secondary structure content in Max p21 (*vide infra*) but appears to have a significant α-helical content in Max p22 (*45*). Furthermore, the (seemingly) more structured N-terminal segment of the p22 isoform was supposed to participate in interactions with the basic region of the protein and in this way to modulate the thermodynamic stability and DNA-binding affinity, as well as to provide the basis for auto-inhibition (*44-46*). It should be noted, however, that neither the exact secondary structure differences, nor their functional importance have been convincingly demonstrated.

1.3.3.2 Structure

All members of the Myc/Max/Mad network contain a basic region-helix-loop-helix-leucine zipper (b-HLH-LZ) structural motif, which is responsible for DNA binding and dimerization, flanked by sequences that are involved in recruiting of effector partner proteins. The known crystal structures of Myc-Max, Mad-Max and Max-Max proteins bound to E-box DNA have revealed a very similar structural organization of the b-HLH-LZ domains in the context of either homodimeric Max-Max or of the heterodimers (*47, 48*).

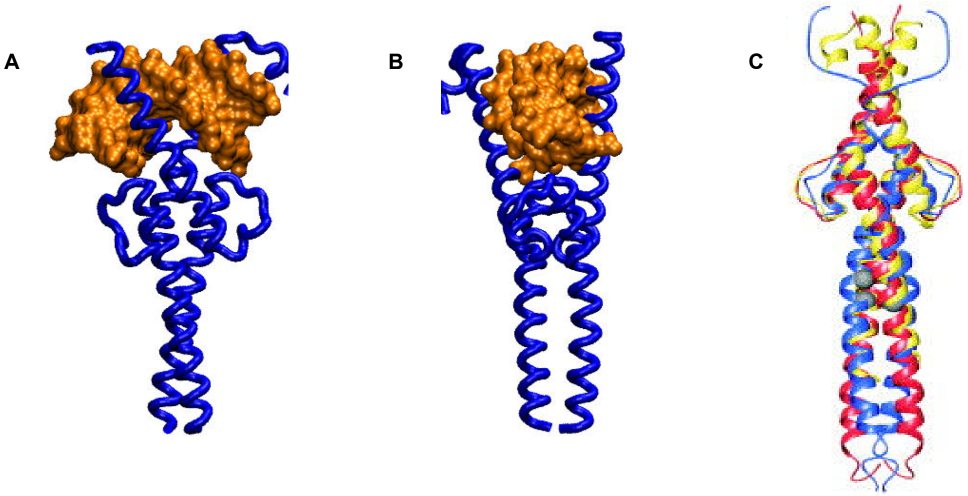

Figure 1.5. Structure of the Max protein bound to E-box DNA (1HLO). In A and B, two approximately orthogonal projections of the complex are shown (rotation by 90° around the long axis of the DNA duplex). The Max protein in tube representation is colored blue; the DNA in surface representation is colored ochre. C, Superposition of the HLH of one of the most representative conformers of the NMR ensemble of free Max (blue) onto the HLH of the two DNA-bound crystal structures of Max (1AN2 in red) and (1HL0 in yellow). Panel C is reproduced from (50).

The HLH domains fold as a parallel four-helix bundle. A parallel two-stranded coiled coil is formed by the adjacent LZ domains. Within the HLH domain, dimerization is facilitated by tight packing of conserved Leu, Ile and Phe side chains. The hydrophobic core is complemented by the non-polar moiety of an Arg side chain located at the start of helix H2. The leucine zipper subdomain has the typical coiled coil architecture. Besides the archetypal Leu-Leu, and Ile-Ile knobs-into-holes packing, the Max leucine zipper exhibits His-His interchain contacts, as well as a Gln-Asn-Gln-Asn tetrad. There is solid evidence that these unusual clusters play an important role to modulate the stability of the Max-Max homodimer as compared to the stability of Max heterodimers with other members of the Myc-Mad-max network (*47, 48*). The very C-terminal part of the leucine zipper is partly unfolded. The structure of Max bound to DNA is shown in Figure 1.5.

The basic region domains form α-helices naturally emerging from the end of helix H1 of HLH and diverging to grip the duplex by fitting into the major groove. It has been long suggested that the basic region segments undergo a conformational transition from a quasi-random coil in the free protein to an α-helix in the protein-DNA complex (*49*). Very recently, the structure of the unbound Max-Max homodimer was determined by NMR, for the first time allowing insights into the structural rearrangements accompanying DNA recognition (*50*). Indeed, the first 14 residues of the basic region are largely unfolded in the absence of DNA. Interestingly, however, the last four residues of the basic region form a highly-populated helical turn, just adjacent to helix H1 of HLH. Since highly conserved basic residues contacting the E-box site are located in this region, it was plausibly suggested that pre-folded α-helical conformers might speed up specific binding. Also, transient helical turns in the N-terminal part of the basic region probably exist, as suggested by the α-helical-like chemical shift of the backbone atoms (*50*). The loop region possesses high flexibility in the unbound state. Little difference was observed in the conformation of the LZ coiled coil domain in the free and in the bound state (Figure 1.5).

1.3.3.3 Biophysical properties

The stability of the full length gene product Max p21 was characterized by thermal melting followed by CD spectroscopy (*51*). No data are available on the stability of isoforms p22 and ΔMax. The Max p21 dimer is highly populated in the low micromolar range. A significant stabilization was achieved by replacing Asn 78 and His 81 from the leucine zipper by valine and leucine, respectively. The finding is in line with earlier studies showing that the same double mutation, or else introduction of a disulfide bond at the C-terminus of the leucine zipper both stabilize the isolated b-HLH-LZ domain (*51*). On the basis of CD melting experiments it was concluded that the N-terminal and C-terminal segments beyond the b-HLH-LZ domain are unstructured. Indeed, high susceptibility to protease digestion is long known (*45*). However, Max p21 is more stable that the “trimmed” b-HLH-LZ domain. The likely reason is the high density of negative charges in the C-terminal extension, helping to relieve the (assumingly) destabilizing electrostatic repulsion caused by the highly positively charged basic regions of the neighboring chains. Although HLH and LZ subdomains are clearly demarcated structurally, it was found that the thermal melting of Max p21 is well described by a two-state transition between folded dimer and unfolded monomer. Stabilization of the leucine zipper by the N78V/H81L mutation produces a pronounced change in the unfolding mechanism. A still dimeric, yet largely non-helical intermediate becomes significantly populated before dissociation of the two polypeptide chains.

The aggregation state of the b-HLH-LZ proteins is still under debate. There are results suggesting that the members of the Myc/Mad/Max family can form dimers of dimers via association patches provided by the leucine zipper domains (*47, 52*). Indeed, tetrameric Myc/Max and Max/Max (but not Mad/Max) species binding two copies of the cognate DNA were observed by X-ray crystallography (*47, 48*). In contrast, solution studies by NMR provided no evidence for the existence of higher-order aggregates of Myc-Max and Max-Max dimers (*50*). The precise conditions and concentrations required for heterotetramer or homotetramer formation *in vitro* an *in vivo* remain unknown.

1.3.3.4 Max-DNA binding

The available biophysical and structural information is consistent with a 1:1 binding model according to which one Max-Max homodimer binds one DNA duplex. A head-to-tail dimer-of-dimers, each dimer binding one duplex was observed by crystallography. However, this observation contradicts studies in solution with protein and DNA in the micromolar concentration range (see Results and Discussion) and the tetramer might be an artifact at the conditions of the crystallographic experiment. There is indirect evidence that monomeric Max could also bind to DNA with very high affinity (*53*). The existence of a monomer-binding pathway is likely of physiological importance at the low concentrations in the nucleus, and has actually been proven for other dimeric transcription factors (*54-58*). However, at the high concentrations required for biophysical experimentation, the dimer-binding pathway clearly dominates. In the case of Max (and other members of the Myc-Mad-Max network) monomer binding has not been detected by direct experiment. So far, only the affinity of truncated Max variants (the isolated b-HLH-LZ domain) to E-box DNA has been determined by electromobility gel-shift assays and fluorescence spectroscopy. Binding is in the low nanomolar range

(K_D 1-50 nM at room temperature). The DNA binding affinity of the full length Max protein at the physiological temperature of 37 °C is not known. It has been suggested that DNA-binding affinity increases for Max variants in which the leucine zipper is stabilized.

X-ray crystallography analyses of Max-E-box complexes have identified the protein side chains contacting the major groove of the type A E-box (5'-CACGTG-3') (*47*). The contacts from the two monomers are quasi-symmetric.

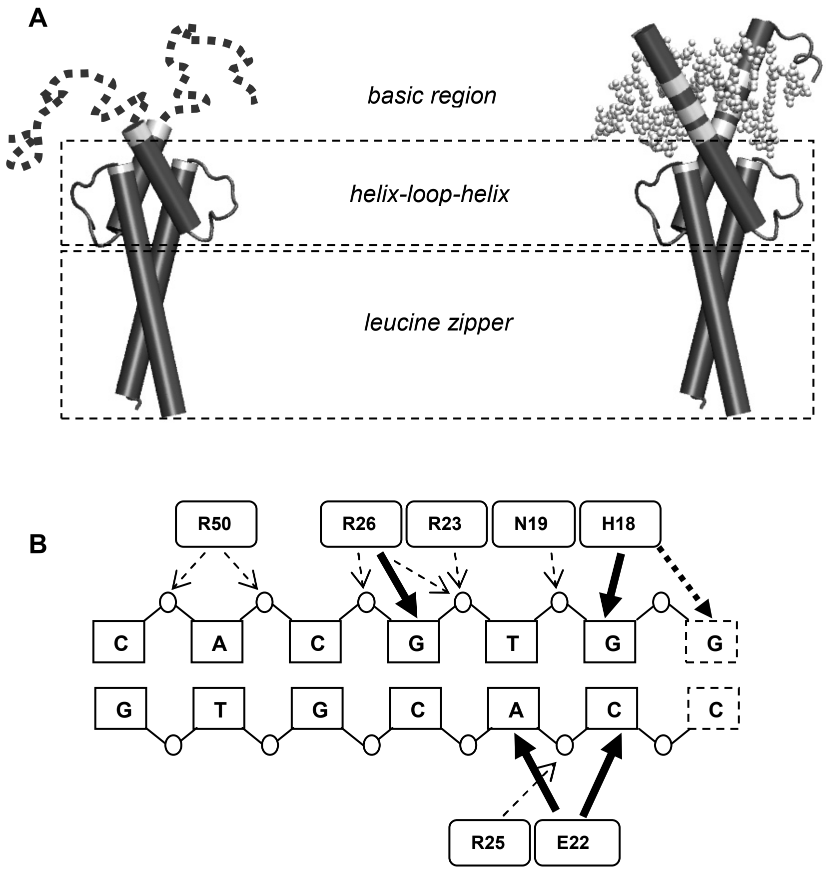

Figure 1.6 Structure of Max p21 and protein-DNA interactions in the Max p21-E-box complex. A, E-box binding does not perturb the structure of HLH and LZ but the basic region folds from a predominantly coil conformation to a regular α-helix fitting into the DNA major groove. In the cartoon representation residues contacting DNA are shown in white. Note that the very C-terminus of the basic region containing the DNA-contacting Arg 25 and Arg 26 are part of an α-helical turn also in the free protein. B, Protein-DNA interactions within the E-box target sequence. The protein-DNA contacts formed by one of the protein monomers are shown as bold arrows (base-specific hydrogen bonds) and thin arrows (contacts to the phosphate backbone.

Structurally, the recognition pattern is essentially identical in the DNA complexes formed by the isolated b-HLH-LZ domain and the full length Max p21 protein (*47, 52*). Seven side chains per monomer are involved in contacts with DNA (Figure 1.6). Six of them are located in the basic region. Base-specific hydrogen bonds are formed by His 18 and Glu 22. These two side chains define the outer E-box cases CA (Glu 22) and GT (His 18). The guanidino group of Arg 26 is hydrogen-bonded to the central G and so determines the identity of the central CG pair. The pattern of base-specific hydrogen bonds is characteristic for site-specific recognition of type A E-box by bHLH proteins. Non-specific electrostatic interactions between side chains from the basic region and the DNA phosphate backbone are made by Asn 19, Arg 23, Arg 25, and Arg 26. In addition, Arg 50 from the loop region contacts DNA through a main-chain and a side-chain contact to backbone phosphates.

1.4 Objectives of the present work

The energetic principles of E-box recognition by full-length b-HLH-LZ proteins of the Myc/Mad/Max family are virtually unknown. So far, only the binding affinity of short constructs encompassing the b-HLH-LZ core domain has been determined. In the present study we use mainly calorimetry to characterize thermodynamically the site-specific binding of the complete Max p21 isoform to a 21 base pair DNA duplex containing the E-box sequence. Specifically, the work intends:

(i) to achieve a complete description of the thermodynamic stability of the Max p21/E-box complex, including the quantification of the enthalpy, entropy and heat capacity changes accompanying binding;

(ii) to provide for the first time a reliable estimate of the dissociation constant at the physiologically relevant temperature of 37 °C;

(iii) to reveal the energetic importance of specific and non-specific protein-DNA contacts by alanine mutagenesis;

(iv) to (re)examine the biophysical properties of the full length gene product of Max p21 isoform with special attention to the influence of non-structured protein segments outside the boundary of b-HLH-LZ on the Max p21 stability and DNA binding properties;

(v) to test by direct binding experiments whether the DNA-binding affinity is influenced by the intrinsic stability of the leucine zipper sub-domain

(vi) to construct and characterize a monomeric, unfolded Max p21 protein as a tool to elucidate the stability of the initial monomer-duplex complex, a crucial (yet elusive) step in the proposed monomer binding pathway of b-HLH-LZ proteins.

2 Materials and Methods

2.1 Sample preparation and buffers

2.1.1 Protein expression, purification and characterization

Figure 2.1 shows the protein constructs used in this study. The full length gene product Max p21 is 150 amino acids long. The short constructs designated $\mathrm{Max_{short}}$ contain the 86 amino acids encompassing the b-HLH-LZ domain plus the C-terminal extension GSGG/C. Max p21VL and $\mathrm{Max_{short}}^{VL}$ bear the double mutation N78V/H81L in the leucine zipper. The C-terminal cysteine in $\mathrm{Max_{short}}^{SS}$ facilitates formation of a covalent link after oxidation. The expression vectors (pET3a plasmid carrying ampiciline resistance) were kindly provided by Dr. Pierre Lavigne, Sherbrooke University, Canada. Details on the construction of the vectors can be found in the original publication (*51*).

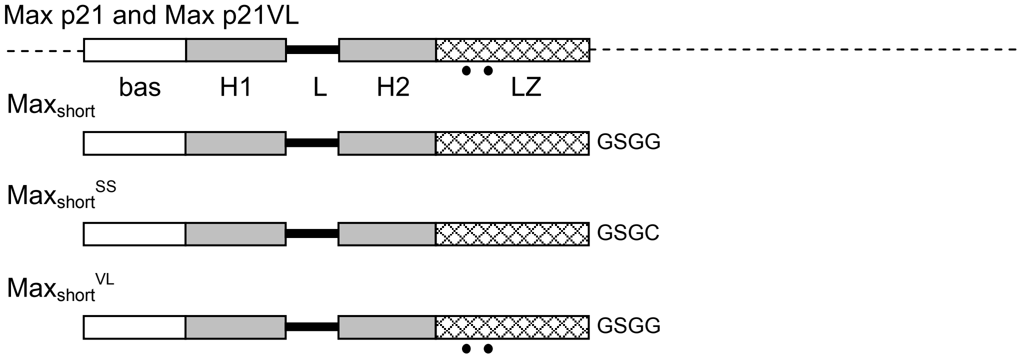

Figure 2.1. Protein variants in this study. Secondary structure elements are indicated.White bar, basic region (bas). Grey bars, α-helices (H1 and H2). Hatched bar, leucine zipper (LZ). Thick line, loop (L). The dashed lines are the unstructured N- and C-terminal segments. The length of the schematic elements is proportional to the corresponding sequence length. The position of the N78V and H81L mutations in Max p21VL are indicated with dots.

Site directed mutagenesis to produce alanine variants in the basic region of Max p21 was carried out using commercially available kits from QIAGEN following the protocol recommended by the manufacturer. The primers designed for mutagenesis PCR are listed in the Appendix. The transformed cells were grown at 37 °C. After induction with 0.6 mM IPTG expression was allowed for additional 3 hours. The typical expression yield was 12 to 28 mg pure protein from 1 L of culture. After expression, the cells were centrifuged, were frozen and stored at −20 °C. The total protein extract was purified by cation-exchange chromatography. Briefly, the cell pellet was resuspended in lysis buffer (50 mM Tris-HCl (pH 7.4), 0.1 M NaCl, 10 mM $MgCl_2$, and 0.1% NP-40). DNAse I was added and the solution was incubated at 37 °C for 1 h. To precipitate DNA and acidic proteins, PEI was added and the solution was centrifuged at 20 000 rpm for 30 min. The supernatant was diluted with 5 vol of buffer A (50 mM sodium acetate (pH 5.0)), and loaded onto a HiTrap SP sepharose HP (Amersham Biotech) column preconditioned with buffer A. The column was washed with 5 vol of buffer B (50 mM sodium acetate (pH 2.8) and 2 M urea). Elution of wild type Max p21 and variants was achieved with a gradient of buffer (50 mM sodium acetate (pH 2.8), 2 M urea, and 3 M NaCl) from 0% to 100%. Fractions containing wild type Max p21 and variants were then desalted on a HiTrap size-exclusion column and lyophilized. Alternatively, the combined fractions after ion exchange chromatography were dialyzed against water overnight, loaded on a semi-preparative C8-HPLC column, and eluted with binary acetonitrile/water gradients (3% to 60%) containing 0.1% and 0.085% trifluoroacetic acid. The material after HPLC was lyophilized. The protein refolded from the lyophilized state was fully native, as judged by CD spectroscopy and displayed DNA binding affinity indistinguishable from that measured with, natively-purified non-lyophilized material. The purity was verified by analytical HPLC with detection at 220 nm and the mass was verified by mass spectrometry. The concentration was measured by UV spectroscopy using ε_{280} = 5960 M^{-1} cm^{-1} calculated from the amino acid sequence. The disulfide linked variant Max_{short}^{SS} was prepared by air oxidation of vigorously stirred concentrated protein solutions for 24-48 h at pH 8.0. The success of disulfide bridge formation was verified by analytical HPLC and mass spectrometry.

2.1.2 Preparation of DNA duplex and protein-DNA complex

Single stranded oligo-nucleotides with the sequences 5’-CCCCCAA**CACGTG**TTGCC TGA-3’ and 5’-TCAGGCAA**CACGTG**TTGGGGG-3’ (the central E-box type A sequence in bold) were purchased from Metabion GmbH (Martinsried). After HPLC purification, they were dissolved in water and the concentration was determined from the UV absorption after complete digestion with phosphodiesterase overnight at

37 °C. The annealing was done by mixing equimolar amounts of both strands in 100 mM potassium chloride. The sample was incubated at 90 °C for 5 minutes and then slowly cooled down during 4 hours. Concentration of the double stranded E-box was determined by UV spectroscopy and proper annealing was confirmed by agarose gel electrophoresis. The protein-DNA complex was formed by step-wise addition of protein to a DNA until the calculated molar 1:1 ratio (protein dimer:DNA duplex) was achieved.

2.1.3 Buffers

All biophysical experiments were performed in a standard buffer containing 100 mM sodium phosphate, pH 6.8. The buffer was supplemented with 100-600 mM KCl. The pH of samples containing urea was adjusted after addition of the denaturant. The urea concentration was determined by measuring the refraction index. All chemicals were of analytical grade and were used without further purification.

2.2 Biophysical methods

2.2.1 Circular dichroism spectroscopy

CD spectroscopy is based on the property of inherently asymmetric chromophores or symmetric chromophores in asymmetric environment to interact differently with right- and left-circularly polarized light. Two related phenomena result. Circularly-polarized light travels through an optically active medium with different velocities due to the different refraction indices for right- and left-circularly polarized light, a phenomenon called *optical rotation* or *circular birefringence*. The variation of optical rotation as a function of wavelength is called *optical rotary dispersion (ORD)*. Right- and left-circularly polarized light is also absorbed to different extents at some wavelengths due to differences in extinction coefficients for the two polarized rays, a phenomenon called *circular dichroism (CD)*. Historically, the unit of circular dichroism is *ellipticity* (θ). Modern CD instruments are capable of millidegree precision. The unit ellipticity persists despite the fact that CD is now measured as the difference in absorbance of right- and left- circularly polarized light as a function of wavelength. Each type obeys the Lambert-Beer law so that the molar ellipticity of a protein sample is $[\theta] = 3298\Delta\varepsilon = 3298(\varepsilon_L - \varepsilon_R)$, where 3298 is a conversion factor.

CD spectroscopy is widely used in modern biophysical studies of proteins. Secondary structure elements and random-coil polypeptides have distinct CD spectra in the far-UV range (below 250 nm). This property facilitates estimates of the secondary structure content in folded proteins and provides a convenient tool to monitor conformational transitions in proteins. To facilitate structural comparisons between proteins of different size, the molar ellipticity or protein samples is usually reported as molar ellipticity per residue, calculated as $\mathrm{MRE} = \frac{\theta}{10.\mathrm{L}.\mathrm{C}.\mathrm{N}}$ (units of deg cm^2 dmol^{-1}), where θ is the observed ellipticity in millidegrees, L is the optical path length in cm, C is the molar concentration, and N is the number of residues.

Experimental procedures

Circular dichroism (CD) spectroscopy experiments were performed on a Jasco J-715 spectropolarimeter equipped with a computer-controlled water thermostat using jacketed quartz cells of 0.2, 0.5 or 1 mm optical path length. Spectra were recorded at scanning speed of 5 nm min^{-1} in 0.2 nm intervals. Three wavelength scans were averaged to obtain the final spectrum. Isothermal urea unfolding experiments were performed at 5, 20 and 25 °C. The samples containing the protein (15 µM dimer equivalents) in different urea concentrations were incubated overnight and the ellipticity at 222 nm was sampled for three minutes after thermal equilibration. The concentration dependence of the ellipticity at 222 nm was studied with 1-100 µM (monomer equivalents) protein solutions at 25 °C. Thermal denaturation was performed by monitoring the change in ellipticity at 222 nm during continuous heating at 1 deg min^{-1} between 5 and 75-90 °C. Reversibility was checked from the recovery of the CD signal after cooling, and was always at least 93%.

Analysis of CD data

Analysis of heat-induced and denaturant-induced curves followed the formalism describing a simple two-state transition between the folded state N and the unfolded state U. At each temperature and denaturant concentration, the observed ellipticity (θ or MRE) is given by:

$$\theta = f_U\theta_U + f_N\theta_N = f_U\theta_U + (1-f_U)\,\theta_N \tag{2.1}$$

where f_U and $f_N = 1 - f_U$ are the fraction of unfolded protein and the fraction of native protein, respectively. θ_U and θ_N refer to the ellipticity of the unfolded and folded state, respectively. θ_U and θ_N are assumed to be linear functions of the temperature or the denaturant concentration of the general form:

$$\theta_\sigma = \theta_{\sigma,0} + \alpha_\sigma T \tag{2.2a}$$

or

$$\theta_\sigma = \theta_{\sigma,0} + \alpha_\sigma[\text{denaturant}] \tag{2.2b}$$

Subscript σ indicates the N or U state. In a two-state unfolding process f_U is easily calculated from the experimental data (be it from thermal unfolding, i.e. $\theta = f(T)$ or from denaturant-induced unfolding at fixed temperature, i.e. $\theta = f([D])$ as:

$$f_U = \frac{\theta - \theta_N}{\theta_U - \theta_N} \tag{2.3}$$

f_u is linked to the equilibrium unfolding constant K_U by the following expressions:

$$f_U = \frac{K_U}{1 + K_U} \tag{2.4a}$$

$$f_U = \frac{-K_U + \sqrt{K_U^2 + 8K_U M_{tot}}}{4M_{tot}} \tag{2.4b}$$

Equation 2.4a applies for unfolding of a monomeric protein ($N \rightleftarrows U$). Equation 2.4b applies for unfolding of a non-covalently linked homodimer, if unfolding is associated with chain dissociation ($N \rightleftarrows 2U$).

In thermal unfolding, the temperature dependence of K_U can be calculated from:

$$K_U(T) = K_U(T_m)\exp\left\{\left[\frac{\Delta H_m}{R}\left(\frac{1}{T_m}-\frac{1}{T}\right)\right]-\frac{\Delta C_p}{RT}\left[T-T_m-T\ln\left(\frac{T}{T_m}\right)\right]\right\} \qquad (2.5)$$

T_m is the temperature where $f_U = f_N = 0.5$; ΔH_m is the unfolding enthalpy at T_m; ΔC_p is the unfolding heat capacity increment; R is the gas constant. Equations 2.1 – 2.5 can be combined to analyze thermal unfolding curves by non-linear curve optimization. Since T_m, ΔH_m and ΔC_p are strongly inter-dependent; the statistical significance of the values extracted from the analysis of a single CD melting trace is low. However, when ΔC_p is small compared to ΔH_m (a condition which holds in protein unfolding), the second term in the curly brackets of equation 2.5 can be neglected. ΔC_p can be obtained directly from differential scanning calorimetry experiments (described in section 2.2.2), or else from plotting ΔH_m *versus* T_m (Kirchhoff's plot, equation 1.12). Knowing T_m, ΔH_m and ΔC_p, the unfolding free energy ΔG_U is calculated from the Gibbs-Helmholtz equation (given here in a slightly different notation than in equation 1.6):

$$\Delta G_U(T) = \Delta H(T_m)\left[1-\frac{T}{T_m}\right]+\Delta C_p\left[T-T_m-T\ln\left(\frac{T}{T_m}\right)\right]-RT\ln K_U(T_m) \qquad (2.6)$$

Note that $K_U(T_m) = 1$ if unfolding is monomolecular. For bimolecular, homomeric transitions, $K_U(T_m) = \frac{2f_U^2 M_{tot}}{1-f_U} = M_{tot}$. In an alternative treatment, the van't Hoff enthalpy of unfolding at T_m (ΔH_m) can be calculated as (*59*):

$$\Delta H_m = (2n+2)RT_m^2\left(\frac{\partial f_U}{\partial T}\right)_{T=T_m} \qquad (2.7)$$

where n is the number of polypeptide chains comprising the native state.

Isothermal chemical denaturation data were analyzed following the linear extrapolation model (LEM) (*60*). The free energy of unfolding in the presence of denaturants, $\Delta G_U(D)$ is given by:

$$\Delta G_U(D) = -RT\ln K_U(D) = \Delta G(H_2O) - m_D[D] \qquad (2.8)$$

where m_D has the units of kJ mol^{-1} M^{-1} and describes the linear dependence of $\Delta G_U(D)$ on [D], the concentration of denaturant. $\Delta G(H_2O)$ is the free energy of unfolding in the absence of denaturant. Non-linear regression analysis of the data with the combined equations 2.1-2.4 and 2.8 yields the values of $\Delta G(H_2O)$ and m_D as the fitting parameters. In practice, faster convergence and higher statistical significance of the fitting procedure was achieved by replacing $\Delta G(H_2O)$ by $[D]_{1/2}$ (the denaturant concentration where $f_U = f_N = 0.5$) as a floating parameter to be optimized. Since at $[D]_{1/2}$

$$\Delta G(H2O) = m_D[D]_{1/2} - RT \ln K_U(D_{1/2}) , \qquad (2.9)$$

equation 2.8 can be re-arranged in the form:

$$\Delta G_U(D) = -RT \ln K_U(D) = -RT \ln K_U(D_{1/2}) + m_D[D]_{1/2} - m_D[D] \qquad (2.10)$$

$K_U(D_{1/2}) = 1$ for monomeric transitions; $K_U(D_{1/2}) = M_{tot}$ for unfolding/dissociation of a homodimeric protein. Once m_D and $[D]_{1/2}$ are known, $\Delta G(H_2O)$ is calculated according to equation 2.9. Data analysis was performed with in-house written scripts for MicroCal Origin 5 and NLREG (Philip Sherod). The stability of a multimeric protein that unfolds *via* dissociation to unstructured monomers can also be evaluated from the concentration dependence of the CD signal. In this case, the data in the form $\theta = f(M_{tot})$ are described by the combined equations 2.1-2.3 and 2.4b.

2.2.2 Differential scanning calorimetry

Among other experimental methods, differential scanning calorimetry (DSC) is the most powerful technique to resolve the energetics of conformational transitions of biological macromolecules. By measuring the temperature dependence of the partial molar heat capacity, DSC directly quantifies all relevant parameters describing the energetics of the native state in respect to the denatured state. Furthermore, statistical thermodynamics analysis provides insights into the thermodynamic mechanism of unfolding (*1*).

Experimental procedures

Differential scanning calorimetry experiments were carried out with a VP-DSC calorimeter (MicroCal Inc.) equipped with twin coin-shaped cells of 0.52 ml volume. Details on the instrument's performance are given elsewhere (*61*). The heating rate was 1 deg min^{-1}. A pressure of 2 atm was applied on the measuring block. The concentrations used were typically in the range 15-60 µM (protein dimer, DNA duplex, or 1:1 complex equivalents) and were always determined after dialysis. Protein, DNA and protein-DNA complex samples were dialyzed overnight against the standard buffer. The same buffer batch was used to establish the instrumental base line. After loading of the sample, the standard protocol consisted of 2-4 heating-cooling scans between 20 and 50 °C to expel microscopic air bubbles from the cells, followed by a heating from 5 to 70-95 °C. With the available instrument cooling scans are not reproducible. The reversibility of the thermal transitions and the repeatability of the melting traces were checked by repeated heating. Reversibility was at least 90 % (typically better), as judged from the value of the calorimetric enthalpy after re-heating. The raw experimental data (differential power; units of J s^{-1}) were normalized for the scanning rate (units of deg s^{-1}) to obtain the C_p/T trace (units of J K^{-1}). After correction for the instrumental buffer-buffer baseline that was recorded according to an identical heating protocol, the C_p/T data were transformed to partial molar heat capacity (MHC; units of J K^{-1} mol^{-1}), or to partial specific heat capacity (SHC; units of J K^{-1} g^{-1}) using the calculated molecular mass, or the partial specific volumes, respectively. (For clarity, I refer to both MHC and SHC as to $C_p(T)$ in the following discussion.) The partial specific volumes were calculated from the amino acid sequence (*62*).

Data analysis

The analysis of heat capacity traces followed the formalism detailed elsewhere (*63*). Briefly, the temperature dependence of the measured heat capacity is expressed as:

$$C_p(T) = C_{p,N} + f_U \Delta C_p + \Delta H_m \frac{df_U}{dT} = C_{p,N} + f_U \Delta C_p + \frac{K_U}{(1+K_U)^2} \frac{\Delta H_m^2}{RT^2} \quad (2.11)$$

K_U (the equilibrium unfolding constant), f_U (the fraction of unfolded protein), ΔH_m (the unfolding enthalpy), and T_m are defined as in section 2.2.1. $\Delta C_p = C_{p,U} - C_{p,N}$ is the unfolding heat capacity change. The heat capacity of the folded protein, $C_{p,N}$, was modeled with a linear function . As shown before, the heat capacity of the unfolded state, $C_{p,U}$, can be calculated with good precision from the amino acid composition of the protein, and is well approximated by a second-order polynomial function of the general form $C_{p,U} = a + bT + cT^2$, where a, b, and c are coefficients. In equation 2.11, it is implicitly assumed that K_U depends on the temperature according to equation 2.5. Depending on whether the unfolding transition is monomolecular or n-molecular, K_U is defined in terms of f_U as in equations 2.4a and 2.4b. Regression analysis according to equation 2.11 returns an optimized value for ΔH_m, which is in fact the geometric mean of the model-independent calorimetric enthalpy, ΔH_{cal}, and the model-dependent van't Hoff enthalpy, ΔH_{vH}, i.e., $\Delta H_m = (\Delta H_{cal} \Delta H_{vH})^{1/2}$. The calorimetric enthalpy is obtained by integration of the $C_p(T)$ trace in respect to the intrinsic heat capacity change function, which is defined by the first two terms on the right-hand side of equation 2.11:

$$\Delta H_{cal} = \int_{T_1}^{T_2} \left(C_{p,U} - C_{p,N}\right) dT = \int_{T_1}^{T_2} \Delta C_p dT \quad (2.12)$$

The van't Hoff enthalpy (effective enthalpy) can be calculated as (*64*):

$$\Delta H_{vH} = \left(\sqrt{n} + 1\right)^2 T_m \sqrt{R\left(C_{p,max} - \frac{\Delta C_p \sqrt{n}}{\sqrt{n} + 1}\right)} \quad (2.14)$$

where $C_{p,max}$ is the heat capacity at T_m. For a macromolecule dissociating to n subunits during thermal melting, the van’t Hoff enthalpy is alternatively evaluated by equation 2.7, in which the fraction unfolded protein (the progress of unfolding) is defined as:

$$f_{U,cal} = \frac{\int_{T_x}^{T_2} \Delta C_p dT}{\int_{T_1}^{T_2} \Delta C_p dT} \tag{2.15}$$

Temperatures T_1 and T_2 demarcate the heat absorption peak, and T_X is any temperature in between. For a two-state transition, $\Delta H_{cal} = \Delta H_{vH}$. If $\Delta H_{cal} > \Delta H_{vH}$, thermal unfolding is not two-state: stable unfolding intermediates accumulate during the transition. If $\Delta H_{cal} < \Delta H_{vH}$ an irreversible process accompanies thermal unfolding.

The calorimetric entropy change of unfolding is calculated by:

$$\Delta S_{cal} = \int_{T_1}^{T_2} \Delta C_p(T)\, d \ln T \tag{2.16}$$

From DSC data, the stability of the protein at any temperature can be calculated by the Gibbs-Helmholtz equation using the calorimetric estimates of ΔH, ΔS and ΔC_p. Data analysis was performed with scripts for Origin 5 provided by the MicroCal Ltd, CpCalc 2.1 (CSC), or in-house written scripts for NLREG (Philip Sherod).

2.2.3 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) enjoys growing popularity as a tool to study binding phenomena involving biologically important molecules. The changes in free energy, enthalpy and entropy can be reliably measured in 2-3 hours. The temperature dependencies of ΔH and ΔS, hence the complete thermodynamic profile of an association reaction, are obtained from only a few of experiments. Thermodynamic parameters are free of artefacts caused by chemical modification or immobilization of binding partners. The binding affinities measured by ITC span the range from 10^4 to near 1×10^9 M^{-1} and many important protein-ligand reactions are open to study.

Experimental procedures

ITC experiments were performed on MCS ITC or VP-ITC mircocalorimeters (Microcal Inc., Northampton, MA). The calorimeters were calibrated according to the manufacturer's protocol. Samples of protein and DNA were dialyzed overnight against the standard working buffer or in 100 mM phosphate buffer containing 250-600 mM KCl, pH 6.8. After dialysis the samples were filtered (Whatman, 0.2 μM) and the concentration was determined by UV absorption at 280 nm (protein) and 260 nm (DNA). Typically, the sample cell (1.4 ml) was loaded with 10–30 μM protein dimer solution, or else with 5-10 μM DNA duplex. The concentration in the injection syringe varied from 50 to 200 μM. The stirring rate was always 300 rpm. The thermal baseline was established with stirring until the drift of the differential power signal was less than 0.01 μcal min^{-1}. A titration experiment comprised typically 25-30 injections, each of 10 μl volume and 10 s duration. At the beginning of the experiment a 2 μl addition was done to eliminate the slightly changed titrant concentration in the tip of the injection syringe caused by diffusion during equilibration.

Data analysis

In power compensation ITC instruments the heat change upon binding is expressed as the electrical power applied by the feedback network to maintain a small temperature difference between the sample and the reference cell (*65*). Each injection of the ligand (L) into the calorimetric cell containing the binding partner (M) produces a deflection of the differential power signal (μcal s^{-1}) from the thermal baseline. Integration of the differential power with respect to time yields the heat change Δq (μcal) between injections *i* and *i-1*:

$$\Delta q = q_i - q_{i-1} \tag{2.17}$$

Δq is caused not only by the shift of the chemical equilibrium in the cell but contains contributions from the heats of dilution and unspecific effects not pertaining to the binding event *per se* (corrections to be discussed below). *Δq* is proportional to the volume of the calorimetric cell, which is known, to the molar enthalpy of binding, which is a constant at constant temperature and pressure, and to the change in the concentration of the complex, Δ[ML] (or bound L, $\Delta[L]_{bound}$), between injections *i* and *i–1*:

$$\Delta q = V_{cell}\Delta H\left([ML]_i - [ML]_{i-1}\right) \tag{2.18}$$

The degree of saturation Y is defined using the equilibrium constant K_A, and the equilibrium concentrations of M and L as:

$$Y = \frac{[ML]}{[M]+[ML]} = \frac{K_A[L]}{1+K_A\,[L]} \tag{2.19}$$

Equation 2.18 can be re-written as:

$$\Delta q = V_{cell}\Delta H[A]_t\left(Y_i - Y_{i-1}\right) \tag{2.20}$$

Most frequently ITC binding data are deconvoluted by non-linear regression analysis according to the following expression, which represents the derivative of equation 2.20 with respect to the total concentration of L added, $[L]_t$ (*65*):

$$\frac{d\Delta q}{d[L]_t} = V_{cell}\Delta H\left(\frac{1}{2} + \frac{1 - \frac{[L]_t}{n[M]_t} - \frac{1}{n[M]_t K_A}}{2\sqrt{\left(\frac{[L]_t}{n[M]_t} + \frac{1}{n[M]_t K_A} + 1\right)^2 - 4\frac{[L]_t}{n[M]_t}}}\right) \tag{2.21}$$

The iterative χ^2-minimization procedure optimizes the values of the binding constant (K_A), the binding enthalpy (ΔH) and the stoichiometry of binding (n; the number of binding sites on M). The total volume increases in the course of the titration. Therefore, the actual concentrations of M and L after injection *i* are $[M]_{t,i} = [M]_{t,0}\left(1-\frac{dV}{V_{cell}}\right)^i$ and $[L]_{t,i} = [L]_{t,0}\left(1-\left(1-\frac{dV}{V_{cell}}\right)^i\right)$, where dV is the injection volume. These corrections have to be introduced in equations 2.18-2.21. The unspecific heats of ligand dilution in the cell, as well as heats produced by the minute differences in the temperatures of the syringe and the cell are corrected for by subtracting the heats measured after full saturation, or else by the heats measured in blank titrations of L into the cell containing plain buffer. Data analysis was performed with the scripts for MicroCal Origin 5 provided by the MicroCal Ltd.

2.2.4 Static light scattering (SLS)

SLS experiments were performed on an Äkta Explorer chromatography system, equipped with a Tricorn Superdex 200 10/300 (Amersham Biosciences) column, the Tri-Angle light scattering detector miniDawn TRI STAR (Wyatt) and the refractory index detector Optilab (Wyatt). The system was washed carefully with at least 1.5 column volumes of water and equilibrated overnight in freshly prepared, filtered (Steritop Millipore 0.22 μM) and degassed standard buffer (section 2.1.3). The protein concentration was between 42 – and 120 μM (dimer equivalents). The column was loaded with 0.1 ml protein solution. The flow rate was 0.5 ml min^{-1}. The data collected was analyzed with the software Astra.

2.2.5 Analytical ultracentrifugation

Sedimentation equilibrium analysis was performed on a Beckman XL-A analytical ultracentrifuge. Protein and DNA solutions were dialyzed overnight against the standard buffer. The concentration of Max p21 was 10 μM and 60 μM (dimer equivalents), corresponding to the typical concentrations used in ITC and DSC experiments, respectively. The concentration of the complex was 1.5 μM. Since the binding constant at 25 °C is 3.8×10^{8} M^{-1} (and is higher at 20 °C), more than 96 % protein/DNA complex was present in the cell. Equilibration was allowed for 40 hours at 20 °C. The data were sampled at rotor speeds of 18000, 20000 and 25000 rpm. Fifty scans (0.001 cm step size) were collected. After correction for the absorption of the reference cell (filled with plain buffer), the last 5 scans were averaged, and a single-species model was fitted to the data using the program WINNONLIN (http://spin6.mcb.uconn.edu/). Protein partial specific volume was calculated from the amino acid sequence. DNA partial specific volume was calculated with the program PSVOL (*66*). The partial specific volume of the complex was obtained from the weight fractions of protein and DNA in complex. In all experiments, the molecular mass best describing the data were all within 10 % of the expected molecular mass. More complicated models, i.e. formation of Max p21 tetramers, free or DNA-bound, did not improve the fitting statistics. (The experiment was kindly carried out by Sasa Bjelic).

2.2.6 Gel filtration chromatography

Size exclusion chromatography was performed at an ETTAN LC System (Amersham Biosciences) equipped with a Superdex 200 10/300 size exclusion column (Amersham Biosciences). After equilibration with 2-3 volumes of working buffer, varying concentrations of protein sample, solved in the same buffer were loaded. Flow rate was 1 ml min^{-1}. The collected data was plotted and analyzed using Mircocal Origin 5.

3 Results and Discussion

The main focus of the presented thesis is the thermodynamic characterization of the site-specific Max p21 interactions with its target E-box DNA. However, studies of the biophysical properties and conformational stability of the protein(s) and DNA duplex are indispensable in providing a coherent description and interpretation of the binding data. Therefore, in Section 3.1 I present and discuss data that are relevant to understand the biophysical behavior and energetics of the Max p21 protein, Max p21 variants and the 21 bp DNA duplex bearing the E-box target sequence. The results of the binding studies are summarized in Section 3.2. For clarity of presentation, some results are omitted from the main text, and are presented and discussed in the Appendix.

3.1 Biophysical characterization of the Max protein

3.1.1 Oligomerization state

All protein variants were over-expressed in BL21(DE3) pLysS *E. coli* as the soluble fraction. The yields per liter of culture varied between 12 and 28 mg after purification. The proteins were purified to homogeneity by combination of ion exchange chromatography, gel filtration chromatography and reversed-phase HPLC. The b-HLH-LZ domain of Max promotes dimerization. As mentioned in Chapter 1, there is evidence that members of the Myc-Mad-Max family can form dimers-of-dimers (*47, 48*). To asses the oligomerization state of the constructs under study, we performed size exclusion chromatography (SEC), static light scattering (SLS) and analytical ultracentrifugation (AUC) experiments. Figure 3.1 shows the SEC profiles collected with differently concentrated samples of $\mathrm{Max_{short}}^{VL}$ and $\mathrm{Max_{short}}^{SS}$. Both proteins elute in a concentration-independent manner with the same retention time, indicating that they have the same size and hydrodynamic properties in the concentration range 50 to 220 µM (monomer equivalents). Differently from $\mathrm{Max_{short}}^{VL}$, which elutes as a single peak, a shoulder/second peak that appears at larger elution volumes is clearly seen in the SEC profiles of $\mathrm{Max_{short}}^{SS}$. In principle, this observation suggests the existence of a higher-order oligomer(s). However, neither the retention time, nor the intensity (integrated absorption) of the second peak, nor the ratio between the intensities of the main and the second peak exhibit dependence on the concentration (compare for example the traces shown in blue and red in Figure 3.1).

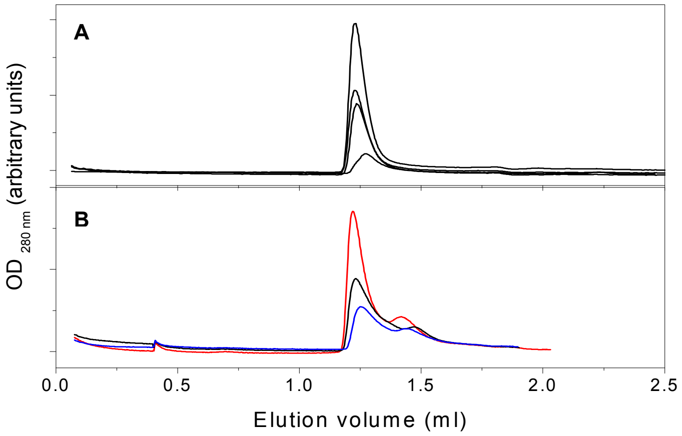

Figure 3.1: Size exclusion chromatography experiments. A, Chromatograms of $\text{Max}_{\text{short}}^{\text{VL}}$ with concentrations between 25 μM and 220 μM. B, Chromatograms of $\text{Max}_{\text{short}}^{\text{SS}}$ with concentrations between 50 μM (in blue) and 100 μM (in red). The concentrations are given in monomer equivalents. All measurements were carried out at room temperature in 100 mM phosphate buffer, 100 mM KCl, pH 6.8.

Any oligomerization process must possess these characteristics. Moreover, DSC experiments $\text{Max}_{\text{short}}^{\text{SS}}$ fail to detect any deviation from the two-state behavior. Therefore, we assume that the second peak in the chromatograms of $\text{Max}_{\text{short}}^{\text{SS}}$ represents some artifact. SEC experiments alone do not provide information about the molecular mass of the proteins (unless molecular mass markers are used at denaturing conditions to calibrate the retention time). SLS experiments were performed to clarify the oligomerization state of $\text{Max}_{\text{short}}^{\text{SS}}$ and $\text{Max}_{\text{short}}^{\text{VL}}$. The results are summarized in Table 3.1.

Table 3.1. Theoretical and measured mass of Max variants.[a]

protein	Concentration (μM monomer)	Calculated mass (Da)	Measured mass (Da)[b]	Measured mass (Da)	Polydispersity index
$\text{Max}_{\text{short}}^{\text{VL}}$	42	9974.15	9974.5	12490 [c]	1.200
	100			20570 [c]	1.010
	120			20800 [c]	1.040
$\text{Max}_{\text{short}}^{\text{SS}}$	62	20116.2	20116.2	21250 [c]	1.007
	100			20660 [c]	1.020
	123			22200 [c]	1.008
Max p21	20	17070.5	17070.4	35600 [d]	
	120			34100 [d]	
Max p21VL		17031.5	17031.5	n.d	
Max $\text{p21}^{\text{L36DM64P}}$		17038.0	17038.1	n.d	

[a] All experiments at 20 °C in 100 mM phosphate, 100 mM KCl, pH 6.8. [b] Determined by ESI mass spectrometry. [c] Determined by SLS. [d] Determined by AUC.

The measured molecular mass of both proteins corresponds perfectly to the theoretical mass of a dimer and the polydispersity index indicates homogeneous species. It should be noted that the actual concentration of the scattering sample is lower than the nominal concentration of the sample applied to the column due to dilution during elution. Therefore, the dimer is highly populated down to at least 20 µM concentration and there is no indication for tetramerization. At the lowest concentration used $\mathrm{Max_{short}}^{VL}$ exists in an equilibrium between monomer and dimer, as judged from the apparent measured mass, which is in between the theoretical masses of the monomer and the dimer, and from the polydispersity index, which deviates substantially from 1. This is not surprising, since the protein is not very stable (see also Figure 3.7). The oligomerization state of the full length Max p21 was characterized by AUC.

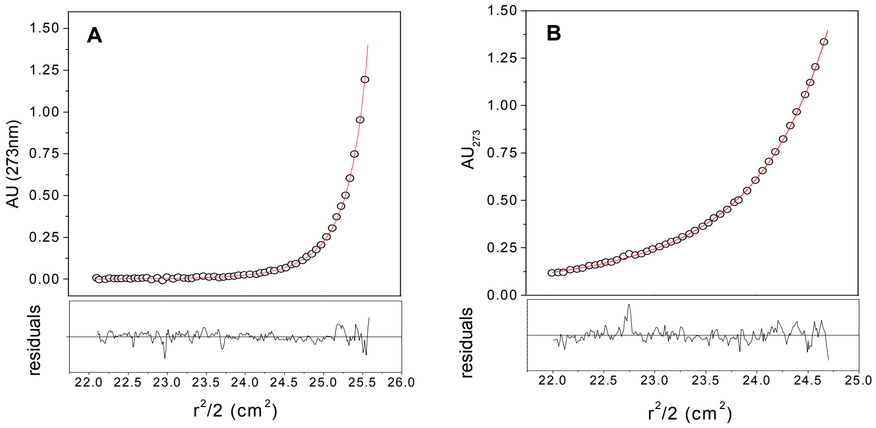

Figure 3.2. Analysis of the oligomerization state of p21 Max and the p21 Max-DNA complex by sedimentation equilibrium analysis. Upper panels, UV absorbance gradients as function of the radial position (circles) and the fits according to a single species model (lines). Lower panels, Residuals showing the difference between the experimental data and the theoretical model. A, p21 Max at 10 µM concentration (dimer equivalents) and 18000 rpm. The fitted mass is 35.6 kDa. B, p21 Max at 60 µM concentration (dimer equivalents) and 25000 rpm. The fitted mass is 37.9 kDa. The calculated mass of the p21 Max protein is 17.07 kDa, i.e. the mass of the Max dimer is 34.1 kDa. C, protein-DNA complex at 1.5 µM concentration at 18000 rpm. The fitted mass is 48.8 kDa; the calculated mass is 47.0 kDa.

As shown in Figure 3.2, the data are perfectly described by a single-species model in the concentration range between 20 and 120 µM. The protein is dimeric (see Table 3.1), and higher-order oligomers are not populated (within the sensitivity of the method). Altogether the data provide solid experimental support that at the solvent conditions and in the concentration regime of both ITC and DSC experiments the Max protein is a dimer.

3.1.2 Secondary structure analysis

Circular dichroism experiments were performed to characterize the overall secondary structure content of Max variants. Figure 3.3 shows the far UV CD spectra of Max p21 and Max$_{short}$, which illustrate the spectral characteristics of the full length protein and its isolated b-HLH-LZ domain, respectively.

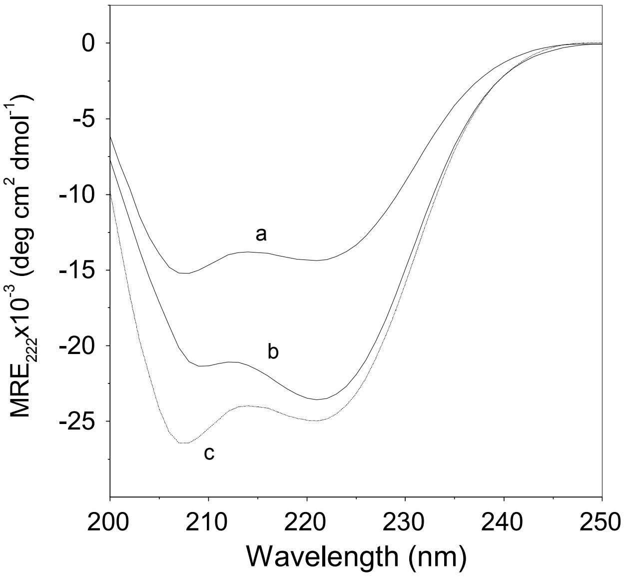

Figure 3.3. Far-UV CD spectra of Max p21 (spectrum *a*) and Max$_{short}$ (spectrum *c*) at 25 °C in standard buffer. Concentrations were 20 µM Max p21 and 100 µM Max$_{short}$ (monomer equivalents). Trace *c* represents MRE$_{222}$ of Max p21 recalculated for 86 amino acid residues per monomer (see the text for details).

Table 3.2. Helical content of Max variants.[a]

protein	MRE$_{222}$	Residues per monomer	Percent α-helix[b]
Max$_{short}$	−23490	86	62
Max$_{short}^{SS}$	−23870	86	63
Max$_{short}^{VL}$	−25140	86	66
Max p21	−14300	150	38
Max p21VL	−15430	150	40

[a] Spectra recorded at 25 °C in 100 mM phosphate, 100 mM KCl, pH 6.8. Protein concentration was 100 µM (monomer equivalents).bCalculated assuming MRE$_{222}$ = −38000 deg cm^2 dmol^{-1} to represent 100 % helix.

Theoretically, a 100% helical protein would have a mean residue ellipticity at 222 nm (MRE$_{222}$) of -38000 deg cm^2 dmol^{-1} (*51*). The measured MRE$_{222}$ and the percentage helical content are listed in Table 3.2.

The short Max variants (b-HLH-LZ) exhibit a ratio MRE$_{222}$/MRE$_{208}$ that is greater than unity, a feature typical for proteins containing interacting α-helices, and are approximately 65 % helical. This helical content is in agreement with the crystal structure (*47*) predicting that in the absence of DNA, i.e. when the basic region is unfolded, about 35 % of the 86 residues are in a random coil conformation (20 % in the basic region, 10 % in the loop, and 5 % in the C-terminal GSCG/C extension). In contrast to $\mathrm{Max_{short}}$ exhibiting CD spectrum of a protein with high helical content, $\mathrm{MRE_{222}}$ of Max p21 is significantly lower. Trace *c* in Figure 3.3 represents the spectrum of Max p21 recalculated by normalization over 86 residues (the length of $\mathrm{Max_{short}}$), instead of 150 residues (the length of Max p21). This allows a more direct evaluation of the spectroscopic signature of the N- and C-terminal segments that are not present in $\mathrm{Max_{short}}$. In the recalculated spectrum, $\mathrm{MRE_{222}}$ of Max p21 is slightly more negative than $\mathrm{MRE_{222}}$ of $\mathrm{Max_{short}}$. The difference is likely due to the fact that the fraction of unfolded (monomeric and non-helical) $\mathrm{Max_{short}}$ is 5-10% at the selected concentration (100 µM), whereas Max p21 is more than 98% dimeric (20 µM). In contrast, the decrease in ellipticity below 210 nm is pronounced and is accompanied by a blue shift. It follows that the ~70 “extra” residues beyond the b-HLH-LZ core domain of Max p21 are devoid of stable secondary structure elements.

3.1.3 Analysis of the specific unfolding enthalpy

The participation of unstructured polypeptide stretches in intra-molecular contacts can not be excluded *a priori* on the basis of spectroscopic data. Scanning calorimetry experiments were performed to check whether the N- and C-terminal extensions of the b-HLH-LZ domain contribute to the unfolding enthalpy. Max p21 exhibits a significantly lower partial specific enthalpy than $\mathrm{Max_{short}}$ (Figure 3.4). When compared at the same temperature (43 °C, selected in between the corresponding melting temperatures to minimize the errors of extrapolation) these enthalpies amount to 10.5 ± 0.5 J $\mathrm{K^{-1}}$ $\mathrm{g^{-1}}$ (Max p21) and 17.1 ± 0.7 10.5 ± 0.5 J $\mathrm{K^{-1}}$ $\mathrm{g^{-1}}$ ($\mathrm{Max_{short}}$). The ratio is thus 0.61, close to the molar mass ratio (0.59). Furthermore, the temperature slopes of the partial specific heat capacities in the native region are very close to each other, ~13 J $\mathrm{K^{-2}}$ $\mathrm{g^{-1}}$.

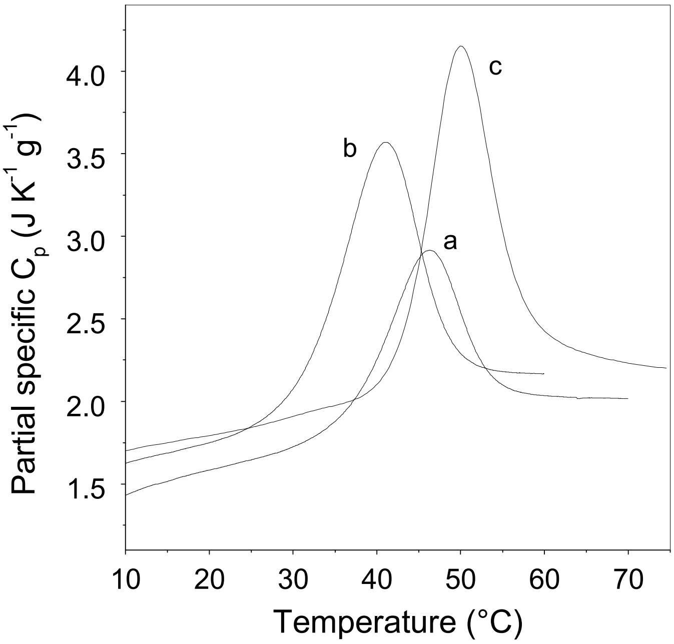

Figure 3.4. Partial specific heat capacity of Max p21 (trace *a*) and Max$_{short}$ (trace *b*) and Max$_{short}^{SS}$ (trace *c*). Concentrations were 125 µM Max p21, 280 µM Max$_{short}$ (monomer equivalents), and 100 µM Max$_{short}^{SS}$. The specific unfolding enthalpy is calculated by integration of the heat absorption peak above a spline function connecting the pre-transitional and post-transitional portion of the thermogram.

Hence, within the limit of detection there are no enthalpically-rich packing interactions involving molecular segments outside the core b-HLH-LZ domain. Altogether, the presented spectroscopic and calorimetric results confirm that the N- and C-terminal domains of Max p21 are unfolded, as it has been concluded previously (*47, 51*). Interestingly, the Max$_{short}^{SS}$ variant, in which the C-terminals of the leucine zipper are linked by a disulfide bond, unfolds with a specific unfolding enthalpy of 20.1 ± 0.2 J K^{-1} g^{-1} (at 43 °C; see Figure 3.4). This observation is in accord with the suggestion that the C-terminal part of the leucine zipper is partly or completely unfolded (*51*). The disulfide link tethers the frying (or unfolded) C-terminuses and assists folding (coiled coiling) along the entire length of the leucine zipper, an effect manifested with heat release.

3.1.4 Unfolding energetics, unfolding mechanism and thermodynamic stability

The thermal stability and unfolding energetics of the Max protein was assessed from thermal unfolding and isothermal urea-induced unfolding. Figure 3.5 shows the results of thermal melting experiments. Although essentially unstructured and not participating in packing interactions with either each other or with the b-HLH-LZ domain, the “tails” confer significant stabilization to Max p21 (Figure 3.5). At the same total protein concentration the mid-point of thermal denaturation of Max p21 is higher by almost 10 °C. The molecular basis of stabilization is unclear at present but it was plausibly hypothesized that stabilization originates from electrostatic effects (*50*). The abundance of negative charges in the N-terminal segment of Max p21 might help to overcome the electrostatic destabilization from charge-charge repulsion between clusters of positive charges at the HLH interface around helix H1 (*50*). The unfolding of $\mathrm{Max_{short}}$, Max p21 and $\mathrm{Max_{short}}^{SS}$ can be described by a simple two-state transition between folded dimer and unfolded monomer. Given the well-defined structural integrity and independence of the HLH and LZ sub-domains, two-state unfolding is surprising. However, several lines of evidence lead to this conclusion. (i) The mid-points of thermal denaturation are the same when evaluated from disruption of secondary structure elements (CD data) and from global disruption of packing interactions (DSC data). (ii) The theoretical functions describing two-state unfolding precisely describe the experimental data (Figure 3.5). (iii) The thermodynamic unfolding parameters are the same when evaluated from CD or DSC data.

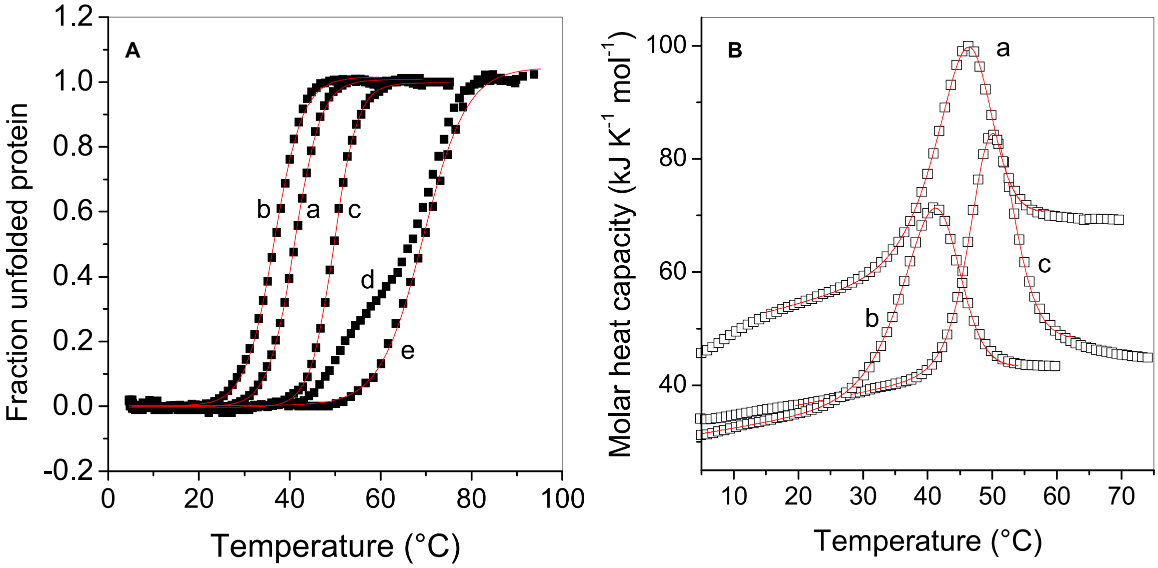

Figure 3.5. Thermal unfolding experiments with Max variants. A, Thermal melting followed by CD spectroscopy. The data are presented as fraction unfolded protein as function of the temperature. The concentration in all experiments was 100 µM (monomer equivalents). The symbols are the experimental data collected with Max p21 (a), $\mathrm{Max_{short}}$ (b), $\mathrm{Max_{short}}^{SS}$ (c), $\mathrm{Max_{short}}^{VL}$ (d), and Max p21VL (e). The continuous lines are best non-linear regression fits according to a two-state model of unfolding according to equations 2.1-2.5. B, Melting followed by DSC. The symbols are the experimentally measured partial specific heat capacity as function of the temperature. The concentration (in monomer equivalents) was 125 µM (Max p21, trace a), 280 µM ($\mathrm{Max_{short}}$, trace b), and 100 µM ($\mathrm{Max_{short}}^{SS}$, trace c). Note that the heat capacity of Max p21 is much larger that that of the shorter constructs, reflecting the large difference in molecular mass. The continuous lines visualize the results of non-linear regression modeling according to equation 2.11. The obtained thermodynamic parameters are listed in Table 3.3.

Table 3.3. Thermodynamic parameters describing unfolding of Max variants.[a]

protein	Concentration (µM monomer)	T_m (°C)	ΔH_{cal}	ΔH_{vH}	$\Delta H_{vH}/\Delta H_{cal}$	ΔS_{cal}	ΔC_p	ΔG (25 °C)
Max_{short}	280	41.1	340	368	1.08	1.01	2.5	37
Max_{short}^{SS}	100	50.1	400	374	0.93	1.24	2.0	28
Max p21	125	46.3	370	400	1.07	1.10	3.6	46

[a] All experiments in standard buffer 100 mM phosphate, 100 mM KCl, pH 6.8. ΔH and ΔG in units of kJ mol^{-1}; ΔS and ΔC_p in units of kJ $K^{-1}mol^{-1}$. Errors are on the order of ± 0.5 °C (T_m), ±20 kJ mol^{-1} (ΔH), ±0.4 kJ $K^{-1}mol^{-1}$ (ΔC_p) and 0.07 kJ $K^{-1}mol^{-1}$ (ΔS).

For example, the unfolding enthalpy of Max p21 at 37 °C is 338 kJ mol^{-1} from CD data and 348 kJ mol^{-1} from DSC data. The unfolding entropy is 1.02 and 0.99 kJ K^{-1} mol^{-1} from CD and DSC, respectively. (iv) Using the temperature of maximal heat absorption (T_{max}), the molar heat capacity at that temperature ($C_{p,max}$), the unfolding heat capacity increment (ΔC_p) and the (assumed) molecularity of unfolding process (n = 2), the effective vant'Hoff enthalpy can be calculated as $\Delta H_{vH} = \left(\sqrt{2}+1\right)T_{max}\sqrt{R\left(C_{p,max} - \frac{\Delta C_p \sqrt{n}}{\sqrt{n+1}}\right)}$ = 400 kJ mol^{-1}. Thus, the ratio $\Delta H_{vH}/\Delta H_{cal}$ is 1.07, in reasonable agreement with a two-state unfolding between folded dimer and unfolded monomer. (v) Finally, the thermodynamic stability calculated from thermal melting experiments and from urea unfolding below 25 °C is identical within error (see below). Therefore, thermal unfolding is highly cooperative.

At the melting temperature of 46.3 °C (125 µM monomer concentration) the calorimetric estimates of the unfolding parameters of Max p21 are ΔH_{cal} = 370 ±20 kJ mol^{-1}, ΔS_{cal} = 1.1 ± 0.1 kJ K^{-1} mol^{-1}, and ΔC_p = 3.6 ± 0.4 kJ K^{-1} mol^{-1}. Table 3.3 summarizes the parameters determined for other Max variants.

The reliability of the parameters listed in Table 3.3 can be verified by studying the concentration dependence of the mid-point of thermal transition (T_m) as function of the total protein concentration (M_{tot}). The general thermodynamic formalism predicts that T_m changes with M_{tot} for an associating system. Using ΔH_{cal}, ΔC_p, and T_m obtained for a given total protein concentration, the temperature dependence of the unfolding constant (dissociation constant, K_D) is calculated according to equation 2.5.

Since at any temperature $K_D = \frac{2f_U^2 M_{tot}}{1 - f_U}$ for a homodimeric protein, at T_m, where f_U = 0.5, K_D = M_{tot}. In other words, the protein is half unfolded at the temperature where K_D = M_{tot}. It has been demonstrated that the maximum of heat absorption (the maximum of the transition heat capacity peak in the thermogram) appears not at the temperature, where f_U = 0.5, as it is the case for monomeric transitions, but at the temperature, where f_U = 0.59 (*64*). It follows that at $T_{Cp,max}$ we have $K_D = 1.698M_{tot}$. Figure 3.6 demonstrates the very good agreement between the experimentally observed $T_{Cp,max}$ and the theoretically calculated function using the parameters from Table 3.3.

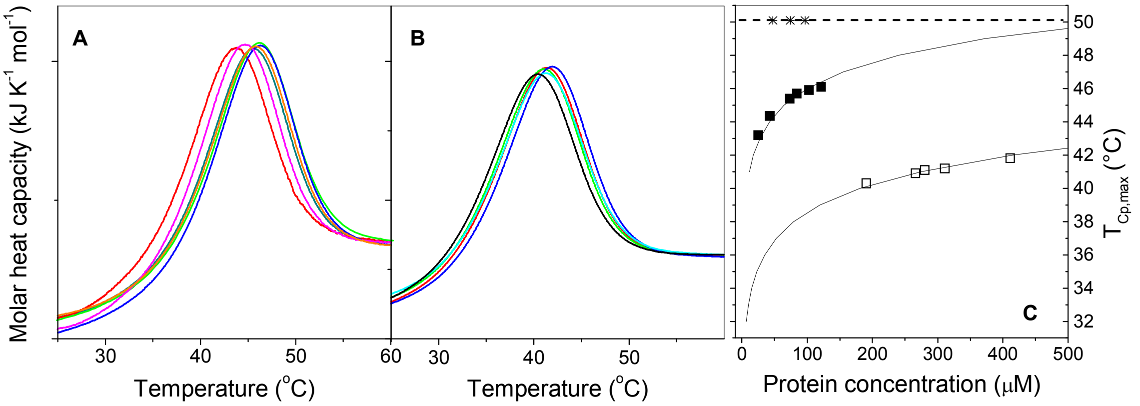

Figure 3.6. Concentration dependence of $T_{Cp,max}$. A, DSC experiments with Max p21. Concentrations wer 26, 43, 74, 84,103,and 122 µM (monomer equivalents; left to right). B, DSC experiments with Max_{short}. Concentrations were 190, 266, 280, 311, and 411 µM (monomer equivalents; left to right). C, Temperature of the maximal heat absorption as function of the protein concentration. Filled symbols, data with Max p21; open symbols, data with Max_{short}. The solid lines are the theoretical functions calculated with the parameters listed in Table 3.3. The data collected with Max_{short}^{SS} are shown with asterisks. This protein is monomer and $T_{Cp,max}$ shows no temperature variation (dashed line).

Stabilization of the LZ sub-domain by covalently linking its C-terminuses significantly increases the thermodynamic stability. Note that ΔG at 25 °C listed in Table 3.3 can not be compared directly, since Max_{short}^{SS} is a monomer (the chains remain linked in the denatured state), while Max_{short} and Max p21 are non-covalent dimers and the unfolding transition is bimolecular. In fact, the so-called effective unfolding free energy (*67*), $\Delta G_{eff} = \frac{\Delta G_{exp}}{n} + \frac{RT}{2}\ln(2M_{tot})$ of Max_{short} and Max p21 at 25 °C are 9 kJ mol^{-1} and 13 kJ mol^{-1}, respectively, more than twice lower than ΔG of Max_{short}^{SS} (28 kJ mol^{-1}).[1] The double N78V/H81L mutation is even more stabilizing (see Figure 3.5). Very interestingly, Max_{short}^{VL} exhibits a very clear change in the unfolding mechanism. Figure 3.7 illustrates that an unfolding intermediate becomes significantly populated. As observed previously, the mid-point of the transition of the native state to the intermediate is concentration-independent, while the mid-point of unfolding of the intermediate to the globally denatured state clearly shifts to higher temperatures when the total protein concentration is higher (51). It follows that the low-temperature transition is monomolecular and the high-temperature transition proceeds with chain dissociation. The exact molecular mechanism of unfolding, i.e. the structural characteristics of the intermediate(s) remain elusive.

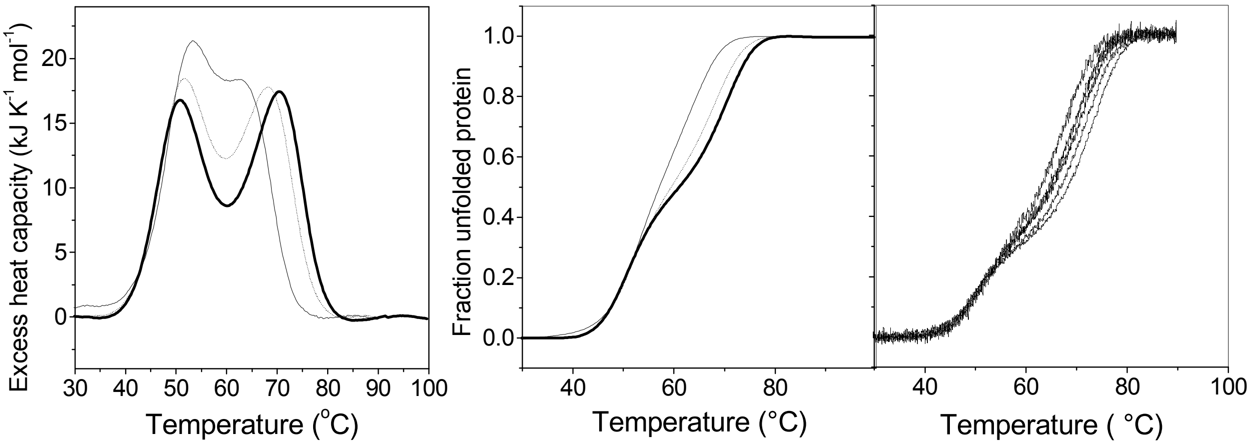

Figure 3.7. Thermal unfolding of Max_{short}^{VL}. A, Excess heat capacity as function of the temperature. The total protein concentration in monomer equivalents was 30 μM (thin line), 79 μM (dotted line) and 130 μM (thick line). B, Progress of thermal unfolding. The fraction unfolded protein was calculated by integration of the excess heat capacity shown in panel A with respect to the temperature and normalization for the total absorbed heat between 40 and 85 °C. The line code is the same as in panel A. C, Fraction unfolded protein calculated from thermal unfolding followed by CD spectroscopy. The total monomer concentration was 35, 59, 76, 94 and 135 μM (left to right). All experiments were performed at scanning rate of 1 deg min^{-1} in 100 mM phosphate, 100 mM KCl, pH 6.8.

On the basis of the loss of helical content in the intermediate as compared to the native state, it was proposed that helix H1 and part of helix H2 are melted in the intermediate, while the remaining part of helix H2 and the leucine zipper are still helical. Our experiments demonstrate that although the helicity of the intermediate is

---

[1] In the equation above, ΔG_{exp} is the experimentally measured free energy change, n is the number of subunits, and M_{tot} is the total concentration in monomer equivalents.

approximately 75% of the helicity of the native state, a significant portion of packing interactions are lost (compare the areas of the two transitions shown in Figure 3.7A).

Similarly to wild type Max, the unstructured N- and C-terminal extensions stabilize also the doubly mutated b-HLH-LZ$^{N78V/H81L}$ domain (compare traces e and d in Figure 3.5). In contrast to $\mathrm{Max}_{short}^{VL}$, Max p21VL unfolds in an apparently monophasic manner, when MRE_{222} is the experimental signal. Formally, the unfolding curve can be reasonably fitted with a two-state transition. However, unfolding is clearly more complicated. One can see that the slope of the $f_U(T)$ function around the mid-point of transition is much lower than for all other variants. Therefore, according to equation 2.7, the unfolding enthalpy of Max p21VL is also lower when evaluated at the same temperature.

This is counter-intuitive because the unfolding ΔC_p is positive (see Table 3.3), meaning an increase of the ΔH at higher temperatures. A much lower unfolding ΔH of Max p21VL in comparison to, say, ΔH of wild type Max p21 would indicate significant structural perturbation linked to disruption of packing interactions as the consequence of the double mutation. This scenario is indeed very unlikely, since Max p21VL is much more stable than Max p21, and the far-UV CD spectra are virtually identical. The apparent paradox is easily resolved by noting that ΔH_{vH} calculated by spectroscopy is an effective parameter reporting on the progress of transition. For Max p21VL ΔH_{vH} at 70 °C is 260 kJ mol^{-1} according to equation 2.7, using the slope $\left(\frac{\partial f_U}{\partial T}\right)_{T=T_m}$ taken from Figure 3.5. At the same time ΔH_{cal} is 450 ± 20 kJ mol^{-1}, calculated by integration of the excess heat capacity profiles in Figure 3.7. The ratio $\Delta H_{vH}/\Delta H_{cal}$ is thus lower than 0.6, a rigorous proof for the presence of an intermediate.[2] It is clear from Figure 3.5 that the intermediate is significantly stabilized in Max p21VL. It unfolds at higher temperatures than in $\mathrm{Max}_{short}^{VL}$ and the N-to-I and I-to-D transitions become highly coupled. As discussed in section 3.1.3, the likely mechanism of stabilization conferred by the N-terminal, acidic unfolded segment of

---

[2] We have no DSC data on Max p21VL unfolding. However, as shown for wild type Max p21, the sequences outside the b-HLH-LZ domain are not involved in packing interactions and have no impact on the magnitude of the unfolding enthalpy. Therefore, it can be safely assumed that ΔH_{cal} of $\mathrm{Max}_{short}^{VL}$ is very similar to ΔH_{cal} of Max p21VL.

Max p21 is neutralization of the destabilizing clusters of positive charges in the vicinity of helix H1, which thought to be unfolded in the intermediate (see above). Taken together the data collected in this study are in line with the idea that destabilization of the four-helix bundle by melting of helices H1 triggers unfolding of Max.

3.1.5 Stability curve of Max

The stability of Max p21 at 5, 20 and 25 °C was assessed from isothermal urea-induced unfolding experiments by following MRE_{222} as function of the denaturant concentration. The data could be modelled with a two-state transition between native and unfolded protein (Figure 3.8).

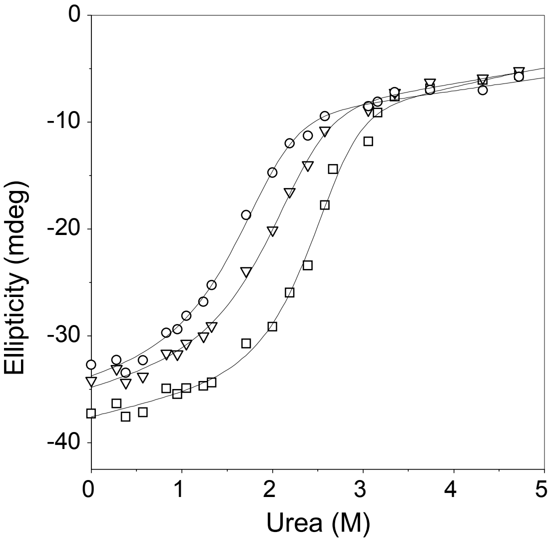

Figure 3.8. Isothermal urea-induced unfolding of Max p21. The symbols are the experimental data collected at 5 °C (squares), 20 °C (triangles) and 25 °C (circles). The continuous lines visualize best non-linear regression fits according to a two-state of unfolding (equations 2.8-2.10). The fitting parameters best describing the data are listed in Table 3.4.

The obtained ΔG values are listed in Table 3.4. As shown in Figure 3.9, the agreement between ΔG_U values thus obtained and data obtained from DSC experiments is excellent. At 37 °C, the dissociation constant is $K_D = 3.6\times10^{-6}$ M and compares well with the only one previous estimate of 7.1×10^{-6} M (*51*).

For the isolated b-HLH-LZ domain K_D at 37 °C was 100×10^{-6} M obtained from CD thermal melting (*51*). Fluorescence anisotropy assay yielded a K_D ~0.7×10^{-6} M at 20 °C (*53*).

Table 3.4. Stability of Max p21 determined by urea denaturation.[a]

Temperature	$[\text{urea}]_{1/2}$	m_D	ΔG
5	2.40	13.8	57.6
20	1.85	10.4	44.6
25	1.40	10.3	41.2

[a] Temperature in °C; [urea]1/2 in M; mD in kJ mol^{-1} M^{-1}; ΔG in kJ mol^{-1}.

For comparison, our estimate for the stability of Max p21 at 20 °C is 0.004×10^{-6} M. These results are in further support to the pronounced stabilization effect of the extensions of the core b-HLH-LZ domain.

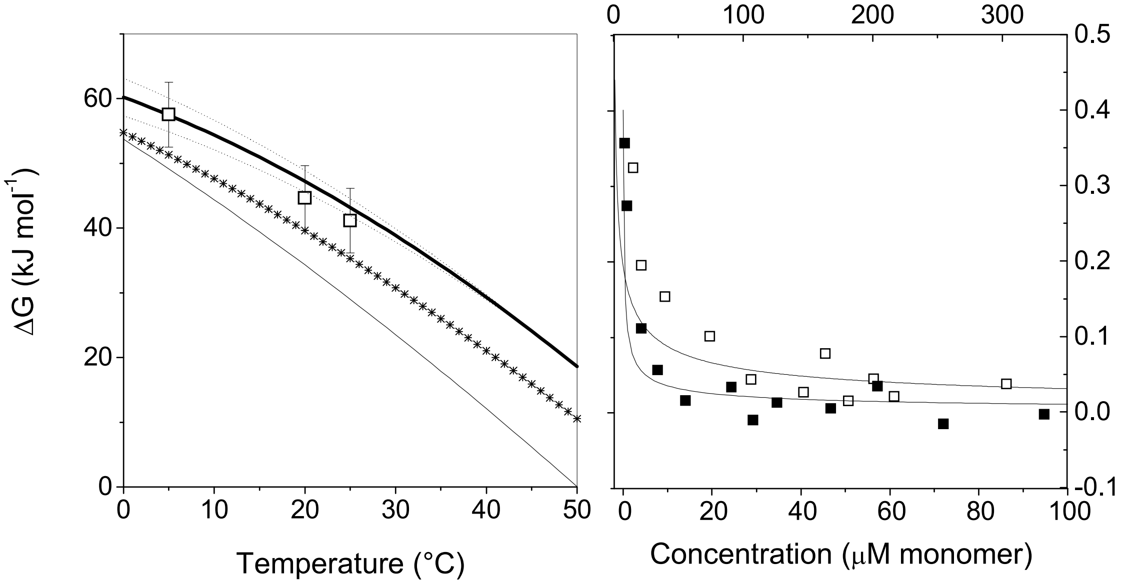

Figure 3.9. Thermodynamic stability of Max variants. A, Stability curves of Max p21 (thick line), $\text{Max}_{\text{short}}$ (asterisks) and $\text{Max}_{\text{short}}^{\text{SS}}$ (thin line). The curves were calculated with the thermodynamic unfolding parameters measured by DSC (Table 3.3). The symbols represent the stability of Max p21 measured by urea-induced unfolding (Table 3.4). The dotted lines associated with the stability curve of Max p21 are the upper and lower error margins introduced by the propagation of the experimental uncertainty at T_m. B, Fraction unfolded Max p21 (solid symbols) and $\text{Max}_{\text{short}}$ (open symbols) at 25 °C as function of the concentration. The solid line is the function calculated as $f_U = \frac{-K_D + \sqrt{K_D^2 + 8K_D M}}{4M}$, where $K_D = \exp\left(-\frac{\Delta G}{RT}\right)$ and M is the concentration in monomer equivalents. The symbols are the experimental data measured by CD spectroscopy (ellipticity as function of the concentration.

Table 3.5 lists the available information on the dissociation constants (equilibrium stability constants) of $\text{Max}_{\text{short}}$ and Max p21. The constructed stability curve (Figure

3.9) was used to simulate the fraction of unfolded protein as function of the temperature and protein concentration.

Table 3.5. Dissociation constants of Max p21 and its b-HLH-LZ domain (Max$_{short}$).[a]

protein	K_D (in µM)	
	20 °C	37°C
Max p21	0.004[b]	3.6 [b] 7.1[c]
Max$_{short}$	0.09 [b] 0.7 [d]	90 [b] 100 [c]

[a] Determined in 100 mM phosphate, 100 mM KCl, pH 6.8. [b] This study. [c] From CD melting (*51*) [d] From fluorescence titration (*53*).

The predicted equilibrium distribution was verified by measuring the change of MRE$_{222}$ for variation of the protein concentration as illustrated in Figure 3.9. The collected information demonstrates consistently that the dimeric state is populated to more than 97% for protein concentrations down to 10 µM between 5 and 25 °C. Binding experiments were designed in this temperature range.

In order to investigate the influence of X-to-Ala mutations in the basic region on the stability of Max p21, we performed CD thermal melting experiments. As expected, mutations of the basic region do not influence the stability of the protein to a measurable extent. The midpoints of thermal denaturation measured by CD spectroscopy for all 7 variants studied here were within ± 1 degree (Figure 3.10).

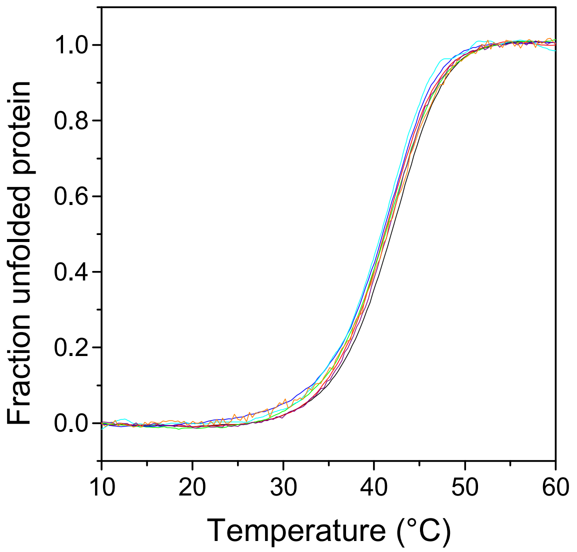

Figure 3.10. Thermal melting of Max p21 basic region X-to-Ala mutants followed by CD spectroscopy. The experimental data (MRE$_{222}$) were normalized and are presented as fraction unfolded protein as function of the temperature. The concentration was 18 µM (monomer equivalents), corresponding roughly to the concentration range of ITC experiments.

3.1.6 Conformational stability of the target E-box DNA

The stability of DNA duplexes is governed by the length and (for the same length) by the G/C content, since G-C Watson-Crick pairs are significantly more stable than A-T pairs (*68*). The G/C content of the DNA duplex used in this study is 62 % and it was expected that it will show significant stability, such that in ITC titrations and in DSC experiments with the protein-DNA complex, there will be no interference from DNA strand dissociation. The stability was probed by CD and DSC thermal unfolding experiments. The CD spectrum of the duplex is typical for short duplexes (Figure 3.11). Upon heating which leads to strand dissociation, characteristic changes occur mainly in two spectral regions. The negative band around 240 nm decreases in intensity. The positive band around 265 nm becomes also less intensive and experiences a red shift (see Figure 3.11 for comparison of the CD spectra at 5 °C and 95 °C). It was shown that both bands are equally suited to follow DNA thermal unfolding (*13*). Thermal unfolding curves were recorded at 264 nm, where the spectral changes between the native duplex and the dissociated strands are maximal.

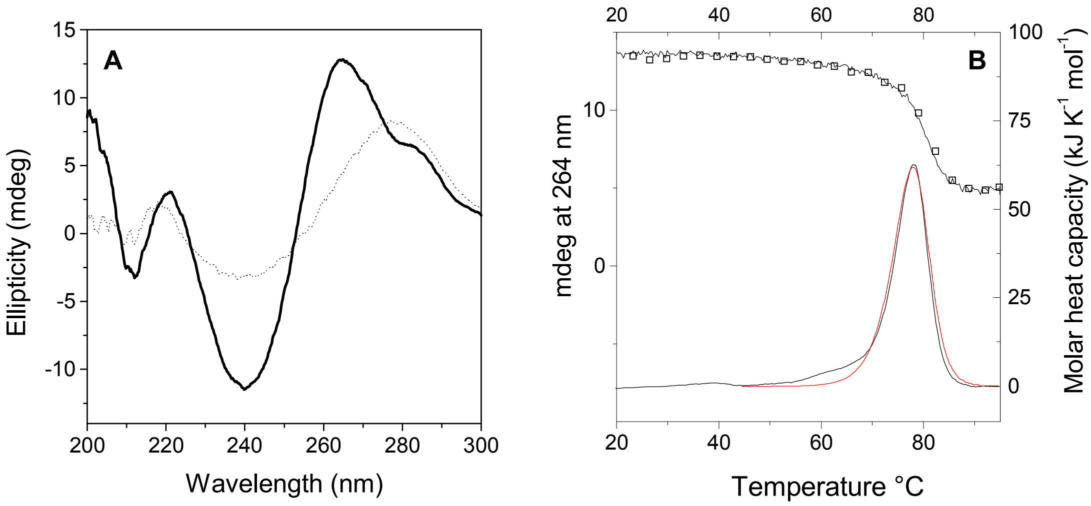

Figure 3.11. Characterization of the thermal stability of the E-box DNA duplex. A, CD spectrum of the duplex at 5 °C (solid line) and 95 °C (dotted line). B, Thermal unfolding experiments. On top, the melitg curve recorded by following the ellipticity at 264 nm is shown (left y-axis). The open symbols is the ellipticity recorded during a cooling scan. Concentration was 100 µM. B, Excess capacity as function of temperature for melting of 17 µM duplex. The red line is the modeled function describing cooperative unfolding with strand dissociation.

As illustrated in Figure 3.11 the duplex undergoes highly cooperative unfolding above 70 °C (the exact melting temperature being dependent on the duplex concentration).

The thermal unfolding was fully reversible: The heating and cooling experiments were superimposable. Incidentally, the full reversibility confirms that the annealing protocol used to make the duplex from the synthetic isolated strands leads to no partially annealed, or else intra-molecular hairpin species. The stability was further characterized by DSC experiments. As seen in Figure 3.11 heating produces a sharp heat capacity peak corresponding to cooperative unfolding with significant heat absorption. The main transition can be modeled with a simple bimolecular process of strand dissociation without intermediate states. The small pre-transition (the shoulder around 60 °C) is likely due to frying of one of the end, probably the one having a terminal A-T pair. The thermodynamic parameters of duplex melting are Tm = 78 °C (17 μM duplex concentration), ΔH_{cal} = 590 kJ mol^{-1}, ΔH_{vH} = 560 kJ mol^{-1}, and ΔC_p 2.3 kJ K^{-1} mol^{-1}. With these parameters, the stability is 100 kJ mol^{-1} at 25 °C. The dissociation constant is thus on the order of 10^{-19} M and, therefore, no strand dissociation takes place in the sub-micromolar concentration.

Altogether, the thermodynamic data presented in this chapter demonstrate that Max p21, Max 21 basic region X-to-Ala mutants, Max p21VL, Max_{short}^{VL}, and Max_{short}^{SS} exist as native dimers between 5 °C and 25 °C down to 10 μM concentration. The binding experiments that are presented in Section 3.2 were performed in this temperature and concentration regime.

3.2 Thermodynamics of complex formation

3.2.1 Experimental design

Protein-DNA binding experiments are typically performed by titrating protein into DNA solution placed in the calorimetric cell. In cases of high affinity binding low reagent concentrations are required to collect data allowing reliable calculation of the association constant and the binding enthalpy from the same binding isotherm. If the protein placed in the injection syringe is composed of non-covalently associated subunits care should be taken that the protein concentration after dilution in the cell is always high enough, so that no dissociation takes place. Unfortunately, the thermodynamic stability of Max p21 is relatively low, as discussed in Section 3.1, and a significant shift of monomer-dimer equilibrium occurs at sub-saturating concentration (0.1-5 μM). In contrast, the stability of the 21-bp E-box DNA duplex is very high. For this reason the ITC experiments were performed by titrating E-box duplex into Max p21. To verify that the choice of titrant and titrand does not influence the results, control experiments of titrating protein into DNA or titrating DNA into protein were performed with $\mathrm{Max_{short}}^{SS}$. The results were identical within error. However, this setup does not overcome the problem that there is redistribution of the population of dimer and monomer during the titration experiment, since the cell contents become progressively diluted and, more importantly, the free p21 Max concentration decreases as the saturation of DNA binding sites increases.

In the following I present a simple model describing how the shape of the ITC isotherm changes when the monomer-dimer equilibrium of the Max p21 protein placed in the calorimetric cell steadily shifts due to formation of the dimer-DNA complexes. For simplicity we assume that all relevant equilibria are faster than the response time of the instrument. There is no kinetic competition between the DNA binding and the folding/refolding. Max p21 monomers are unable to bind DNA (see section 3.2.6) At any given total concentrations of protein dimer (P_T) and DNA duplex (D_T), the equilibrium concentrations of the protein-DNA complex (PD) can be calculated from the only physically meaningful root of the following quadratic equation:

$$\mathrm{K_A PD^2 + \left(P_T + D_T + \frac{1}{K_A}\right)PD + K_A P_T D_T = 0} \qquad (3.1)$$

where K_A is the association constant. The free concentration (P_F, in dimer equivalents) in the cell (after correction for the dilution upon injectant addition) is

$$\mathrm{P_F = P_T - PD} \qquad (3.2)$$

After each injection the fraction of unfolded dimer is:

$$\mathrm{f_{u,i} = \frac{-K_D + \sqrt{K_D{}^2 + 4K_D P_F}}{2P_F}} \qquad (3.3)$$

K_D is the dissociation constant characterizing the monomer-dimer equilibrium of max p21. The concentration of unfolded free protein dimer after completion of the i-th injection is:

$$P_{F,i}^{unf} = P_{F,i} f_{u,i} \quad (3.4)$$

The change in monomer-dimer equilibrium between consecutive injections is described by the change of P_F^{unf} between injections i-1 and i:

$$\Delta P_{F,i}^{unf} = P_{F,i}^{unf} - P_{F,i-j}^{unf} = P_F \Delta f_U = P_F \left(f_{U,i} - f_{U,i-j} \right) \quad (3.5)$$

The heat associated with the shift of the dimer-monomer equilibrium is therefore:

$$\Delta q = q_i - q_{i-j} = \Delta P_F^{unf} \Delta H V_{cell} \quad (3.6)$$

where ΔH is the molar unfolding enthalpy of the dimer at the temperature of the ITC experiment and V_{cell} is the volume of the calorimetric cell. Figure 3.12 below illustrates how the heat associated with the monomer-dimer equilibrium shift contributes to the total measured reaction heat.

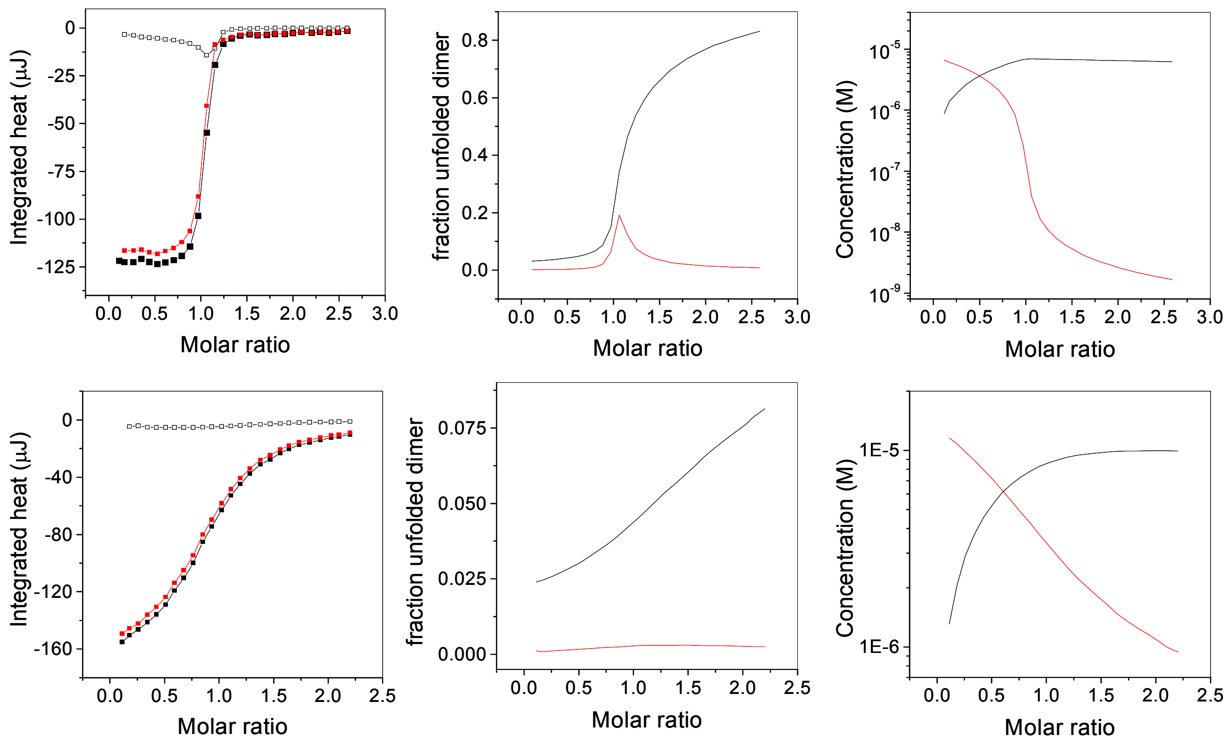

Figure 3.12. Distortion of the heat signal measured in when DNA is titrated to Max p21. The upper panels A-C represent high-affinity binding regime (K_A= 3.8 x 10^8 M^{-1}, 7.9 μM Max dimer placed in the cell, 90 μM DNA in the injection syringe, c-value 3000) Lower panels D-F represent low-affinity binding rgime (K_A=7.5x10^5 M^{-1}, 13 μM Max dimer placed in the cell, 130 μM DNA in the injection syringe, c-value 10) In the calculations the experimentally determined dissociation constant $K_{D(25°C)}$ is 2.7x 10^{-8} M and unfolding enthalpy ΔH(25°C)=295 kJ mol^{-1}. In panels A-D the black fitted symbols are the experimentally measured integral heats at each injection. The open symbols are the heats associated with the shift of monomer-dimer equilibrium. The corrected heats (filled symbols minus open symbols) are shown in red. In panels B and D, the black line is the fraction free unfolded dimer at each injection, $f_{U,i}$. The red line is the difference Δf_U=$f_{U,i}$-$f_{U,i-j}$. In panels C and F, the increase of protein-DNA complex (black line) and the simultaneous decrease of free monomeric Max (red line) are shown.

The figure presents the maximal possible distortion at the highest temperature (25°C) and with concentration in the lowest range used in the ITC experiments. The effect is smaller at lower temperatures since the protein is more stable and, therefore both Δf_U and the unfolding ΔH of Max p21 get smaller as the temperature decreases. The maximal systematic overestimation of the molar binding enthalpy is 6 kJ mol^{-1}, as compared to the standard deviation of 6 kJ mol^{-1} from triplicate experiments made with different stock solutions (including also the error in concentration determination). The "distorted" binding constants are within the usual error range ±10-50%. We demonstrate below that the reported binding constants from ITC are in very good correspondence with data derived from independent experiments. Nevertheless, the reported parameters should be considered as apparent, although they appear quite reliable. It is important to note that the energetic differences observed for basic region mutant complexes are unaffected, since all Max p21 mutant proteins exhibit the same thermodynamic stability. It should also be stressed that the apparently negligible influence of the monomer-dimer equilibrium is the fortuitous consequence of the properties of the considered system. In cases where the intrinsic binding enthalpy is low or the enthalpy of protein unfolding is large, or the protein is less stable, or any combination of these conditions takes place, the distortion of the binding isotherm could be pronounced.

3.2.2 Thermodynamic profile of the wild type Max p21/E-box complex

Examples of binding isotherms measured by ITC are shown in Figure 3.13. Addition of aliquots of E-box DNA duplex to Max p21 produces heat effects, which saturate as the molar ratio of duplex to p21 Max dimer increases.

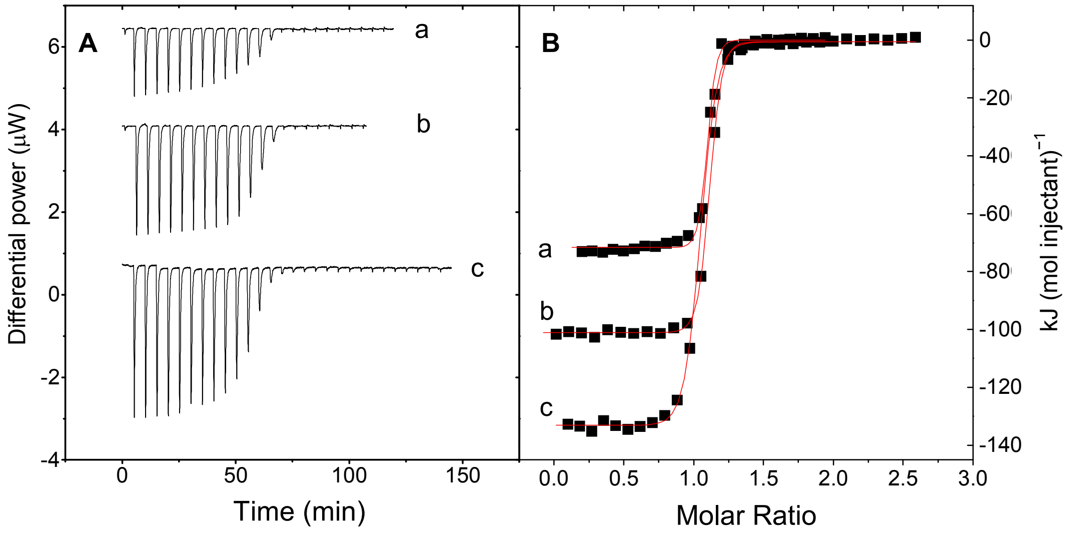

Figure 3.13 Representative ITC binding experiments with Max p21 and E-box DNA. A, Raw experimental output. The traces are shifted on the y-axis for clarity. B, Binding isotherms. The symbols represent the integrated heats released in each addition of DNA to protein, after normalization for the molar concentration and subtraction of unspecific heats. The lines are the results of non-linear regression analysis according to a 1:1 binding model. The experiments were performed in 100 mM phosphate, 100 mM KCl, pH 6.8 at 8 °C (a), 15 °C (b) and 25 °C (c).

The stoichiometry in the data set was 0.98 ± 0.07 p21 Max dimer per DNA duplex (mean ± SD of more than 50 experiments, including experiments with Max p21 variants), in agreement with a 1:1 protein-DNA complex seen in the crystal structures of Max p21/E-box, c-Myc/Max/E-box and Mad/Max/E-box complexes (*48*).

There are suggestions that the members of the Myc/Mad/Max family can form dimers of dimers via association patches provided by the leucine zipper domains (*47, 48, 69*). Indeed, tetrameric Myc/Max and Max/Max (but not Mad/Max) species binding two copies of the cognate DNA were observed by X-ray crystallography (*48*). However, ultracentrifugation experiments performed with concentrations typically used in ITC experiments at our experimental conditions failed to detect the presence of tetramers (Figure 3.14).

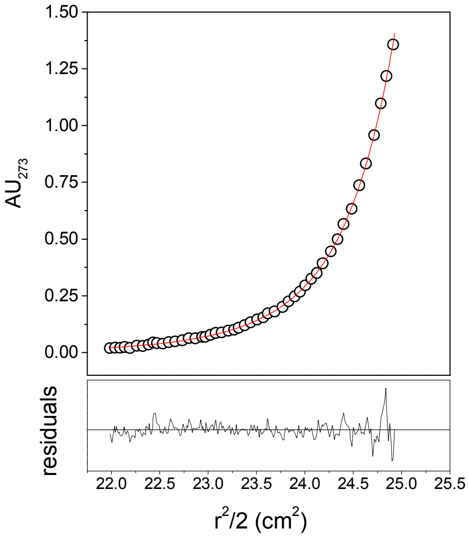

Figure 3.14. Analysis of the ologomerization state of the Max p21-DNA complex by sedimentation equilibrium analysis. Upper panel, UV absorbance gradients as function of the radial position (circles) and the fits according to a single species model (line). Lower panel, Residuals showing the difference between the experimental data and the theoretical model. Concentration was 1.5 µM protein-DNA complex. Rotor speed was 18000 rpm. The fitted mass is 48.8 kDa; the calculated mass is 47.0 kDa.

Therefore, the measured thermodynamic parameters truly reflect the energy changes promoting formation of the Max p21 dimer to the E-box site.

Between 5 and 25 °C association is driven by enthalpy and is opposed by entropy. In this temperature range binding is too strong and only lower estimates of the association constant, K_A, can be obtained with reactant concentration in the low micromolar range. However, the binding enthalpy change, ΔH_A, and its temperature dependence, $\Delta C_{p,A}$, are defined with high precision. The same ΔH_A was measured in phosphate and Tris buffers having different heats of protonation. It follows that there is no (net) proton release/uptake accompanying binding. In order to quantify the strength of binding we explored the well known sensitivity of protein-DNA complexes to the concentration of cations (*70*). Experiments were performed in increasing concentrations of KCl at 25 °C. In control experiments, we confirmed the earlier observation that Max p21 is fully native in the salt concentration regime used here (*69*) and data not shown. As illustrated in Figure 3.15 the binding isotherms become increasingly shallower at higher salt concentration because binding gets weaker (by a factor 55 from 350 mM to 650 mM cation concentration), and reliable K_A can be calculated.

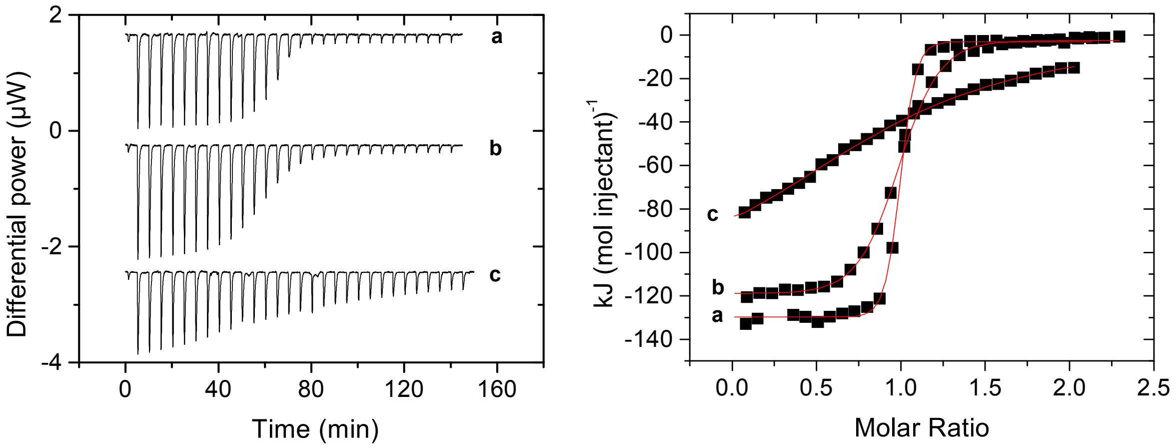

Figure 3.15. Binding isotherms measured by ITC for Max p21 binding to E-box DNA. Titrations were performed at 25 °C in 100 mM phosphate buffer, pH 6.8 supplemented with 150 mM (a), 350 mM (b) and 600 mM (c) KCl. A, Raw experimental data. B, Integrated and normalized heats after correction for unspecific effects. Solid lines are best non-linear fits to a 1:1 binding model.

As in many documented cases, in the absence of specific anion effects $\ln K_A$ depends linearly on the logarithm of cation concentration ($\ln[C^+]$) in the range of salt concentrations where the water activity is negligibly perturbed by the presence of salt. The experimentally measured $\ln K_A = f(\ln[C^+])$ function is shown in Figure 3.20 and will be discussed in Section 3.2.4 in a more general context. Regarding the strength of binding, the important information to be extracted is the extrapolation from high-salt conditions to our standard buffer conditions. This extrapolation yields $K_A(25\ °C) = (3.8 \pm 0.2)\times10^{8}\ M^{-1}$ ($K_D = (2.7 \pm 0.1)\times10^{-9}$ M).

Whereas changes of K_A with temperature, pH and pressure have a rigorous thermodynamic basis, this is not the case for the effect of salt. To verify the reliability of the linear extrapolation DSC experiments were designed to obtain an independent estimate of K_A. To this end, DSC melting traces were recorded with Max p21, E-box DNA duplex and the 1:1 Max p21/E-box complex. The mid-points of thermal unfolding of Max p21 and E-box DNA are separated by ~35 °C (Figure 3.16).

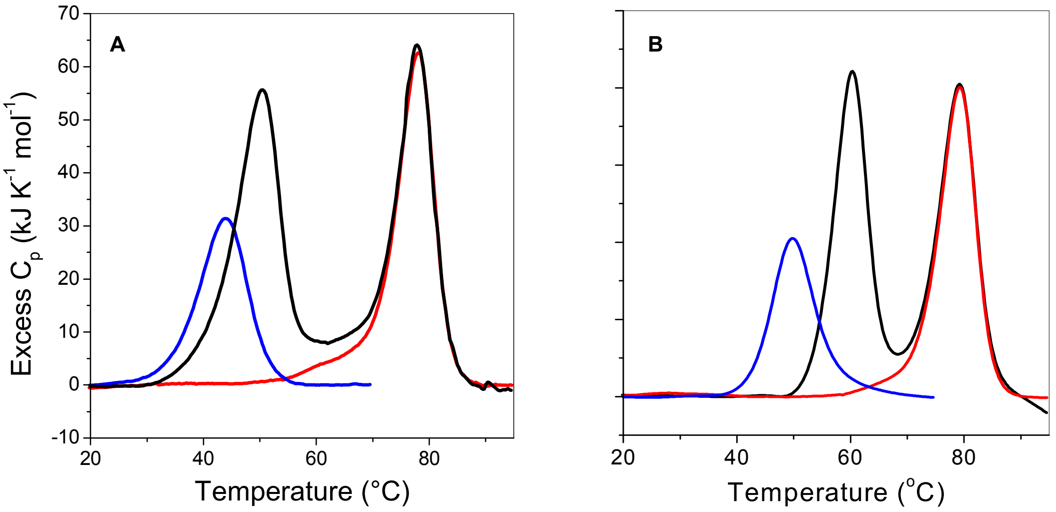

Figure 3.16. DSC melting profiles of Max p21 (A) and Max$_{short}^{SS}$ (B) free and in complex with E-box DNA. The melting profiles of the free protein, free DNA and 1:1 protein/DNA complex are colored in blue, red, and black, respectively. The concentration of all species was 17 µM (A) or 34 µM (B) (protein dimer, DNA duplex and complex). All experiments were performed in the standard working buffer following an identical heating protocol. For clarity, the data are presented as excess heat capacities *versus* temperature and are normalized at 20 °C.

Melting of the 1:1 complex produces two peaks. The first one overlaps partially with the thermogram of isolated Max p21 and corresponds to cooperative protein-DNA complex melting, which is intimately coupled to Max p21 unfolding. The ~7 °C shift in the temperature of maximum heat absorption relative to the maximum of the Max p21 peak reflects the stabilization of the protein caused by protein-DNA contacts. Melting of the complex releases free DNA duplex, which melts at much higher temperature (the second peak of the trace shown in black in Figure 3.16) Using the known unfolding enthalpy (ΔH_P) and unfolding heat capacity change $\Delta C_{p,P}$ of the Max p21 at the melting temperature (T_P) in the absence of DNA (data taken from Table 3.3) the association constant K_A at the temperature corresponding to half-completion of complex melting (T_C) can be calculated according to the following equation (*71*):

$$K_A(T_C)=\frac{\exp\left\{-\frac{\Delta H_P}{R}\left(\frac{1}{T_C}-\frac{1}{T_P}\right)+\frac{\Delta C_{p,P}}{R}\left(\ln\frac{T_C}{T_P}+\frac{T_P}{T_C}-1\right)\right\}-1}{[D]_{T_C}} \qquad (3.7)$$

The term $[D]_{TC}$ in the denominator of equation 3.7 is the concentration of free DNA duplex at T_C. Since binding is strong, the population of the 1:1 complex is higher than 0.98 at the onset of heat absorption (~30 °C) and, therefore, $[D_{T_C}]=\frac{[D]_{tot}}{2}$. The calculated $K_A(50°C)$ is $(1.9 \pm 0.3)\times10^6$ M^{-1}. This number compares well with $K_A(50\ °C)$

= (1.3 ± 0.5)×10^6 M^{-1} obtained by using the parameters directly measured by ITC at 25 °C according to:

$$K_A(T_C) = K_A(25°C)\exp\left\{-\frac{\Delta H_A(25°C)}{R}\left(\frac{1}{T_C}-\frac{1}{298.15}\right)+\frac{\Delta C_{p,A}}{R}\left(\ln\frac{T_C}{298.15}+\frac{298.15}{T_C}-1\right)\right\}$$

(3.8)

The data from ITC (ΔH_A and $\Delta C_{p,A}$) and DSC (K_A(50 °C)) can be combined to yield an estimate of K_A at 25 °C:

$$K_A(25°C) = K_A(50°C)\exp\left\{-\frac{\Delta H_A}{R}\left(\frac{1}{298.15}-\frac{1}{323.6}\right)+\frac{\Delta C_{p,A}}{R}\left(\ln\frac{298.15}{323.6}-\frac{298.15}{323.6}+1\right)\right\}$$

(3.9)

The resulting number is (6.9 ± 2)×10^8 M^{-1} as compared to (3.8 ± 0.2)×10^8 M^{-1} obtained by linear extrapolation of data collected at high salt conditions. The reliability of the procedure was verified on the example of the Max_{short}^{SS}/E-box complex (data in Figure 3.16B). This variant complex is stronger than the Max p21/E-box complex (see below). Again, the binding constants estimated either from extrapolation of lnK_A-*versus*-[C^+] plots or from application of equations 3.7-3.9 were in very good agreement (see Table 3.8). The consistency of the binding constants in an extended temperature interval allows a reliable estimate of the stability of the Max p21/ E-box at 37 °C: (4 ± 1)×10^7 M^{-1} (K_D = (25 ± 6)×10^{-9} M). To our knowledge, this is the first direct estimate of the binding affinity of the complete gene product of Max p21 to E-box at the physiologically relevant temperature. Previously, the affinity of the b-HLH-LZ to E-box containing DNA duplexes was estimated to be in the range 1 and 30 nM at temperatures between 20 and 37 °C and at similar buffer conditions (*53*). It follows that the protein segments outside the boundary of b-HLH-LZ are not involved in any significant interactions with the DNA.

The thermodynamic parameters describing formation of the Max p21/E-box complex at 25 ° C are ΔG_A = –48.9 ± 2.0 kJ mol^{-1}, ΔH_A = –137 ± 8 kJ mol^{-1}, ΔS_A = –0.295 ± 0.025 kJ K^{-1} mol^{-1} ($T\Delta S_A$ = –88 ± 8 kJ mol^{-1}), and $\Delta C_{p,A}$ = –3.8 ± 0.2 kJ K^{-1} mol^{-1}. The thermodynamic profile is shown in Figure 3.17. As already discussed, since binding is tight, ITC experiments provide reliable information only on ΔH_A and $\Delta C_{p,A}$.

The apparent ΔG_A values are underestimates. Consequently, the apparent $T\Delta S_A = \Delta H_A - \Delta G_A$ overestimate the negative entropic contribution. In the Figure, the "genuine" ΔG_A and $T\Delta S_A$ functions are presented (see the Figure caption for details).

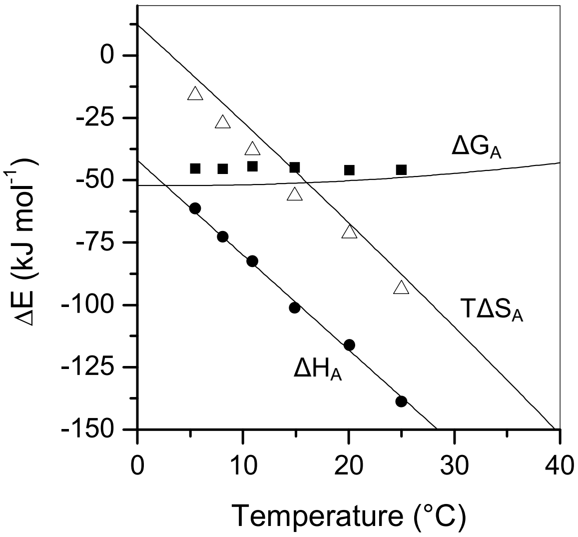

Figure 3.17. Thermodynamic profile of the Max p21/E-box complex. The symbols are the apparent measured ΔG_A (squares), ΔH_A (circles)) and $T\Delta S_A$ (triangles). Since binding is very tight, ITC experiments yield very reliable data on ΔH_A and $\Delta C_{p,A}$. The apparent ΔG_A are underestimates of the genuine stability of the complex. As the consequence the apparent $T\Delta S_A = \Delta H_A - \Delta G_A$ overestimates the genuine negative entropy change. The solid line associated with the ΔH_A data is the best linear fit. The line associated with ΔG_A data was calculated with the parameters listed in Table 3.8 using equation 3.8. The line associated with $T\Delta S_A$ data was calculated as $-[T(d\Delta G_A/dT)]$.

Both ΔH_A and $T\Delta S_A$ are large numbers, which can not be interpreted as the enthalpy accumulated in intermolecular bonds, or as the entropy change associated with these bonds. Rather, the total binding parameters contain the energetic expenditures for the conformational transition of the basic region from the largely unfolded, free state to its α-helical, bound state. Nevertheless, in the following we ask whether the measured energetic terms could be rationalized in structural terms.

Enthalpy change

Considering ΔH_A, the "genuine" binding enthalpy can be estimated by subtracting the enthalpy of α-helix formation from the total observed enthalpy change. Fourteen residues per Max chain undergo coil-to-helix transition (*72*). Depending on the chemical identity of the side chain involved the enthalpy of helix formation was estimated as –3.8 kJ mol residue^{-1} for arginine, lysine, alanine, aspartic acid, glutamic acid, leucine and histidine, and as –2.5 kJ mol residue^{-1} for serine and asparagines (*73*). The weight of evidence suggests –4 kJ mol residue^{-1} as the upper limit for a typical hydrogen bond (*74*). Combining these numbers, the "genuine" enthalpy stabilizing the Max p21-E-box complex can be estimated as –25 to –40 kJ mol^{-1}. Assuming that formation of 10 hydrogen bonds, as seen in the crystal structure, represents the major enthalpic contribution to binding, the mean enthalpic content of such bonds is –2.5 to –4.0 kJ mol^{-1}, in agreement with the estimated enthalpic content of hydrogen bonding in water cited just above. Very similar picture was recently suggested for DNA recognition by the b-ZIP domain of GCN4, on the basis of direct measurement of the enthalpic contribution for folding of the similarly long basic region (*75*).

Entropy change

The large unfavorable entropy change indicates that the loss of conformational entropy dominates the entropic benefits from dehydration of molecular surface and the polyelectrolyte effect. As discussed in Chapter 1 the total entropy change can be formally parsed in terms of different contributions. Table 3.6 lists individual entropic contributions, which were calculated using data from the literature. The calculation is indeed very crude, yet it might give some idea about the balance of entropic factors involved in complex formation.

Table 3.6. Contributions to the experimentally observed total entropy change for Max p21 binding to E-box DNA.

contribution	ΔS at 25 °C (J K^{-1} mol^{-1})	Reference
Immobilization of 22 side chains becoming buried by >20 Å^2 in the complex	−300	(a)
Immobilization of the backbone of 28 residues	−420	(b)
Dehydration of molecular surface	+310	(c)
Polyelectrolyte effect	+120	(d)
Loss of rotational/translational degrees of freedom	−35	(e)
Total predicted	**−325**	
Experiment	**−295**	

(a) Doig et al (1995) Protein Sci. 4, 2247-2251; (b) DAquino et al., (1996), Proteins. 25, 143-156; (c) Luque et al., (1998) Meth. Enzymol. 295, 100-127; (d) Record et al., (1991) Meth. Enzymol. 208, 291-343; (e) Baker et al., (1998) Meth. Enzymol. 295, 294-315

Heat capacity change

Dehydration of the protein-DNA interface would predict a $\Delta C_{p,A}$ of –0.4 J K^{-1} mol^{-1}, ten times lower than the experimental value. Dragan et al. convincingly demonstrated that the similar discrepancy observed for GCN4 b-ZIP binding to DNA is caused by the temperature-induced changes in the structural content (and the extent of DNA-induced refolding for that matter) of the basic region in the temperature range of ITC experiments (*75*). In principle, the effect can be quantified by comparing the temperature dependence of the heat capacities measured for the associated and dissociated state of the protein-DNA complex (*21*). Unfortunately, the limited solubility of Max p21 precluded precise heat capacity measurements in sufficiently concentrated solutions. Nevertheless, the sum of the heat capacities of the isolated components is significantly higher than the heat capacity of the Max p21/DNA complex and the difference strongly depends on the temperature. The contribution of partial refolding and temperature-induced enthalpy fluctuations to the apparent measured $\Delta C_{p,A}$ can be estimated as follows.

Panel A in Figure 3.18 presents the temperature dependence of the partial molar heat capacities of free Max p21 (C_p^{MAX}; purple), free E-box duplex (C_p^{DNA}; red) and the 1:1 protein-DNA complex (C_p^{COMP}; green). Due to the limited accessible range of concentrations, the absolute heat capacities are not precisely known. For clarity, the traces are shifted to zero on the y-axis at 5 °C. The blue line represents the hypothetical heat capacity of the system in the dissociated state, calculated as the algebraic sum $C_p^{SUM} = C_p^{MAX} + C_p^{DNA}$ (blue = purple + red). At any temperature, the heat capacity change of association, taking the dissociated state as the reference state, is $\Delta C_{p,A} = C_p^{COMP} - C_p^{SUM} = C_p^{DIFF}$. The function C_p^{DIFF} is shown in Figure 3.18A with the black line. The integral $\lambda = \int_{T_R}^{T} C_p^{DIFF} dT$ is the enthalpic contribution to ΔH_A arising from the fact that the heat capacities of the associated and dissociated states do not change in parallel upon temperature increase. The absolute value of $\lambda(T)$ can be evaluated at any arbitrary temperature only if $\lambda(T_R)$ is known.

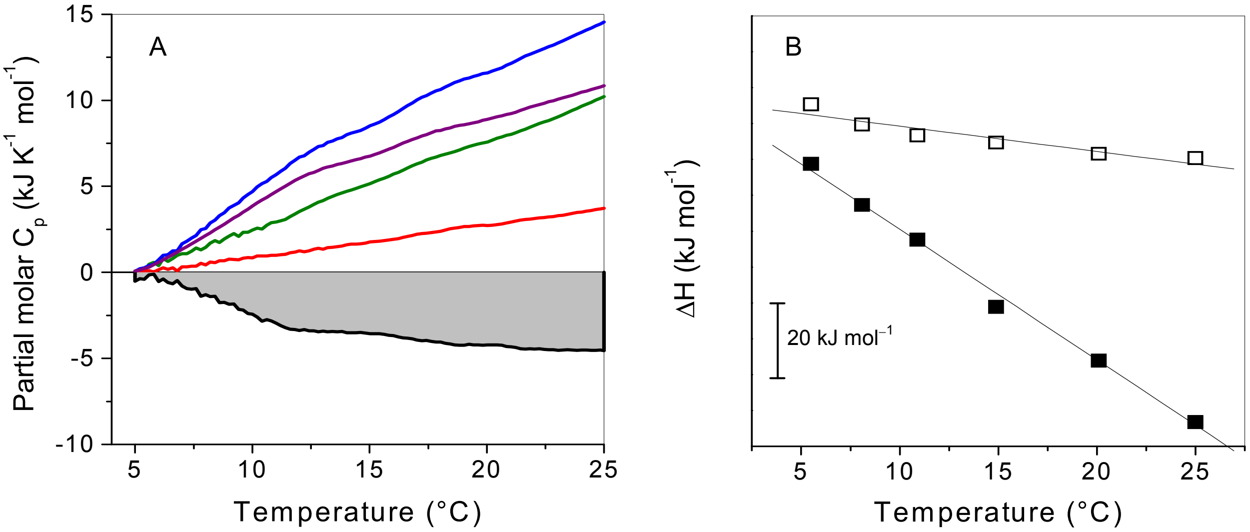

Figure 3.18. Estimation of the contribution of partial refolding and temperature-induced enthalpy fluctuations to the apparent measured $\Delta C_{p,A}$. A, The temperature dependence of the partial molar heat capacities of free Max p21 (C_p^{MAX}; purple), free E-box duplex (C_p^{DNA}; red) and the 1:1 protein-DNA complex (C_p^{COMP}; green). Blue, Hypothetical heat capacity of the system Black, $C_p^{DIFF} = C_p^{COMP} - C_p^{SUM}$. Shaded area ,λ(T)^{rel} between 5 °C and 25 °C. See the text for details. B, Experimentally measured ΔH_A (filled symbols) and the function $\Delta H_A^{corrected} = \Delta H_A - \lambda(T)^{rel}$ (open symbols), arbitrarily shifted on the y-axis. The temperature dependence of ΔH_A is –3.8 kJ K^{-1} mol^{-1} and that of $\Delta H_A^{corrected}$ is –0.9 kJ K^{-1} mol^{-1}.

The latter function could in fact be regarded as the enthalpy of association at some temperature where there are no contributions to ΔH_A other than the enthalpy of formation of intermolecular bonds between “rigid-bodies” and the associated enthalpy of dehydration of groups, and changes of vibration. In the present case $\lambda(T_R)$ is not known. However, the relative values of λ(T) can be calculated with reference to an arbitrarily selected T_R where $\lambda(T_R) = 0$. The shaded area in Figure 3.18A represents λ(T)^{rel} between 5 °C and 25 °C. Figure 3.18B shows the experimentally measured ΔH_A (filled symbols) and the function $\Delta H_A^{corrected} = \Delta H_A - \lambda(T)^{rel}$ (open symbols), arbitrarily shifted on the y-axis. The temperature dependence of ΔH_A is –3.8 kJ K^{-1} mol^{-1} and that of $\Delta H_A^{corrected}$ is –0.9 kJ K^{-1} mol^{-1}. The latter value could be taken as a rough estimate for the binding heat capacity change in the absence of basic region refolding and heat capacity effects associated with the unstructured segments of Max p21. The presented procedure is described in more details in (*13*). In conclusion, from the available data we estimate that the temperature-induced shift of the coil-to-helix equilibrium of the basic region contributes ~ −3 kJ mol^{-1} K^{-1} to the apparent ΔC_p. Hence, ΔC_p arising from intermolecular interactions and dehydration of the surface is ~–1 kJ mol^{-1} K^{-1}, or even smaller (see above).

3.2.3 E-box-binding affinity of Max variants with stabilized leucine zipper

In section 3.1 I demonstrated that stabilization of the leucine zipper sub-domain Max of b-HLH-LZ significantly increases the thermodynamic stability of the protein, both in the context of the full length Max p21 and of the isolated b-HLH-LZ (see Figure 3.5). It has been observed earlier by CD spectroscopy that the mid-point of thermal dissociation of the Max-DNA complex depends on the stability of the Max protein itself (*51*).

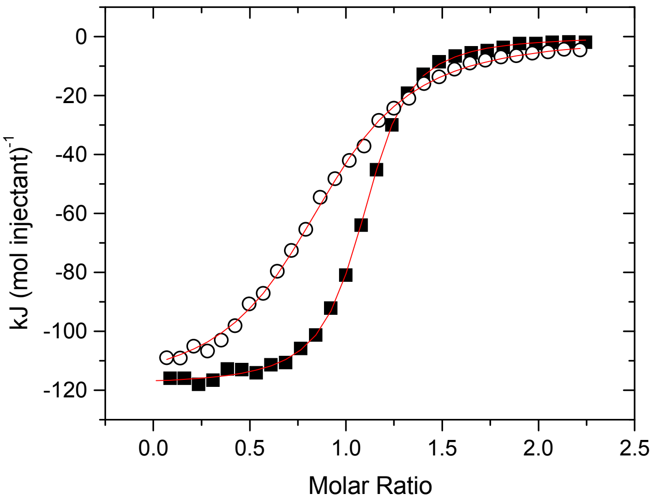

Figure 3.19. Binding isotherms measured by ITC for binding of Max p21 (open symbols) and Max p21VL (solid symbols) to E-box DNA. Symbols, integrated and normalized heats. Lines, non-linear regression fits according to a 1:1 binding model. Since the experiments were performed with similar concentrations, the steeper the titration curve, the stronger the binding. The experiments were performed in 100 mM phosphate, 400 mM KCl, pH 6.8

Table 3.7. Number of cations released upon formation of the Max p21/DNA and estimation of the electrostatic and non-electrostatic contribution to ΔG.[a]

protein	Z^b	$\Delta G^{nel\ c}$	$\Delta G^{el\ d}$
Max p21	10.3 ± 0.1	–26.5 ± 0.2	–22.4 ± 0.3
Max p21VL	13.8 ± 0.1	–25.9 ± 0.2	–29.9 ± 0.3
$\text{Max}_{\text{short}}^{\text{SS}}$	13.7 ± 0.1	–25.5 ± 0.2	–29.7 ± 0.3
Max p21^{N19A}	9.9 ± 0.1	–31.2 ± 0.2	–21.6 ± 0.3

[a] At 25 °C, pH 6.8, ΔG values in kJ mol^{-1}. [b] Calculated as $Z = -\frac{d\ln K_A / d[C^+]}{0.64}$

[c] Non-electrostatic contribution to ΔG_A estimated from –RTlnKA at 1 M cation concentration. [d] Electrostatic contribution to ΔG_A estimated from $\Delta G^{el} = \Delta G_A – \Delta G^{nel}$.

The data has been interpreted as indication that stabilization of the leucine zipper results in an increase of the DNA binding affinity. The present work provides direct evidence supporting this conjecture from binding experiments at room temperature. As an example, Figure 3.19 compares the binding isotherms of Max p21 and Max p21VL in the presence of 400 mM KCl.

As seen in Tables 3.7 and 3.8, either introduction of an S-S bridge at the C-terminus of the leucine zipper of the core b-HLH-LZ domain, or the double N78V/H81L mutation in the context of the full length Max p21 increases K_A by a factor of 12-15. From the data taken at face value from Table 3.8 it would appear that stabilization of the leucine zipper favours binding entropically, especially when the chains of b-HLH-LZ are covalently linked in the $\text{Max}_{\text{short}}^{\text{SS}}$ variant. Still, the observed decrease in ΔH_A off-setting the entropic benefit is not easy to understand. In Max p21VL better binding is accomplished by simultaneous action of enthalpic and entropic factors, yet the enthalpy-entropy balance remains elusive, since the differences are clearly within the experimental error margins. As we explain in the next section, there are reasons to hypothesize that DNA contacts from the loop region and/or helix H1 are created or become stronger in the stabilized versions of Max b-HLH-LZ. Whatever the molecular and thermodynamic mechanisms of the increased affinity might be, the result provides biophysical basis for possible biomedical applications. Since Myc-Max heterodimers bind the E-box target with the same or higher affinity as Max homodimers (*53*), stabilized versions of the latter will more efficiently compete with Myc-Max for E-box binding, and are, therefore, expected to reduce the biological effects of Myc-Max (cell growth and proliferation).

3.2.4 Evaluation of the number of protein-DNA backbone contacts from thermodynamic linkage

The linear dependence between $\ln K_A$ and $\ln[C^+]$ allows estimation of the number of cations released from the phosphate backbone upon protein binding, i.e. the number of protein-phosphate backbone ionic contacts (*70*). Furthermore, the electrostatic and non-electrostatic components of the total binding free energy can be estimated (*76*). For the equilibrium process

$$\mathrm{P} + \mathrm{D} + n\mathrm{C}^+ \rightleftharpoons \mathrm{PD} + m\mathrm{C}^+$$

the following equation holds (*44*):

$$\ln K_A = \ln K_A^{ne} - Z\psi \ln\left[C^+\right] \tag{3.10}$$

P, D and PD represent the equilibrium concentrations of protein DNA and protein-DNA complex, respectively; $[C^+]$ is the molar concentration of cations; *n* and *m* are the stoichiometric coefficients. Z = (n–m) represents the number of protein-DNA ionic contacts, and $\psi = 0.64$ is the number of released cations per phosphate group (*77*). From the slope of the plot according to equation 3.10, Z = 10.3. Therefore, the estimated number of ionic contacts bridging Max p21 to the E-box phosphate backbone is 5 per Max p21 chain. Crystallographic analysis identified the existence of 6 such contacts (*47*). Given the quality of the experimental data, the discrepancy is unlikely caused by experimental uncertainties. Rather, we presume that one of the suggested contacts is not realized in solution at the selected conditions. Incidentally, the same sensitivity on the cation concentration as the wild type complex was observed for the Max $p21^{N19A}$ mutant (Z = 9.9; see Figure 3.20). Since Asn 19 is one of the side chains implicated in electrostatic interaction with the phosphate backbone, it appears plausible to assume that the contact seen in the X-ray structure is not populated in solution. Interestingly, the closest distance of Asn 19 (via OD1) to the DNA backbone is larger than 3.8 Å in the structure of the short Max b-HLH-LZ/DNA complex (1AN2; (*52*)). In fact, the replacement of the Asn19 side chain by alanine is stabilizing the complex, which is difficult to comprehend if favourable polar interactions with DNA occur at that site. Surprisingly, the more stable complexes formed by Max p21VL and Max_{short}^{SS} are also more sensitive to the concentration of cations. Formation of both complexes is accompanied by the release of 7 cations per binding site (Z = 13.8 for Max p21VL and Z = 13.7 for Max_{short}^{SS}; see Table 3.7 and Figure 3.20). It follows that some additional electrostatic contacts in these variants increase the DNA-binding affinity and these new contacts are restricted to side chains within the b-HLH-LZ domain.

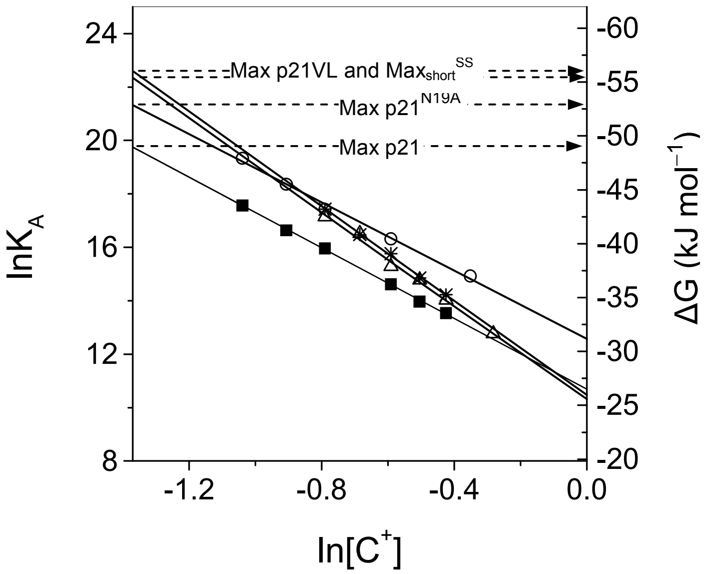

Figure 3.20. Influence of salt concentration on the affinity of Max p21 variants to E-box DNA. The binding constants were measured by ITC at 25 °C in increasing concentrations of KCl. Filled squares, wild type Max p21. Circles, Max p21^{N19A}. Triangles, Max$_{short}^{SS}$. Asterisks, Max p21VL. The lines are linear fits according to equation 3.10. The slopes of the straight lines equal the product Zψ as defined in equation 3.10. Extrapolation to the left y-axis yields $K_A = e^{lnKA}$ at the standard buffer conditions (0.154 mM cation concentration), which are listed in Table 3.8. Extrapolation to the right y-axis yields the non-electrostatic contribution to ΔG_A (ΔG^{ne}) estimated at 1 M cation concentration. The dashed arrows indicate the total ΔG od binding measured experimentally (Table 3.8)

The conclusion is supported by analysis of the total binding free energy in terms of electrostatic and non-electrostatic components. The non-electrostatic component of ΔG_A can be roughly estimated from $-RT\ln K_A$ according to equation 3.10 in the limit of $[C^+] = 1$ M, since the second term of eq 1 is zero. Although it is not possible to estimate to what extent electrostatic effects are attenuated at this salt concentration, it appears that they contribute almost half of the binding free energy measured for Max p21 binding at 25 °C ($\Delta G^{el} \sim -22$ kJ mol^{-1}; see Table 3.7 and Figure 3.20). It turns out that ΔG^{el} is significantly larger in the case of Max p21VL and Max$_{short}^{SS}$ binding ($\Delta G^{el} \sim -30$ kJ mol^{-1}), In fact, the observed increase in ΔG^{el} completely explains the stronger binding of Max p21VL and Max$_{short}^{SS}$, since the non-electrostatic component of ΔG_A, $\Delta G^{nel} = \Delta G_A - \Delta G^{el}$, is identical within error for Max p21, Max p21VL and Max$_{short}^{SS}$. In contrast, ΔG^{el} calculated for Max p21^{N19A} binding is the same as for the wild-type binding, the affinity increase being governed by non-electrostatic effects in this case (ΔG^{nel}; Table 3.7).

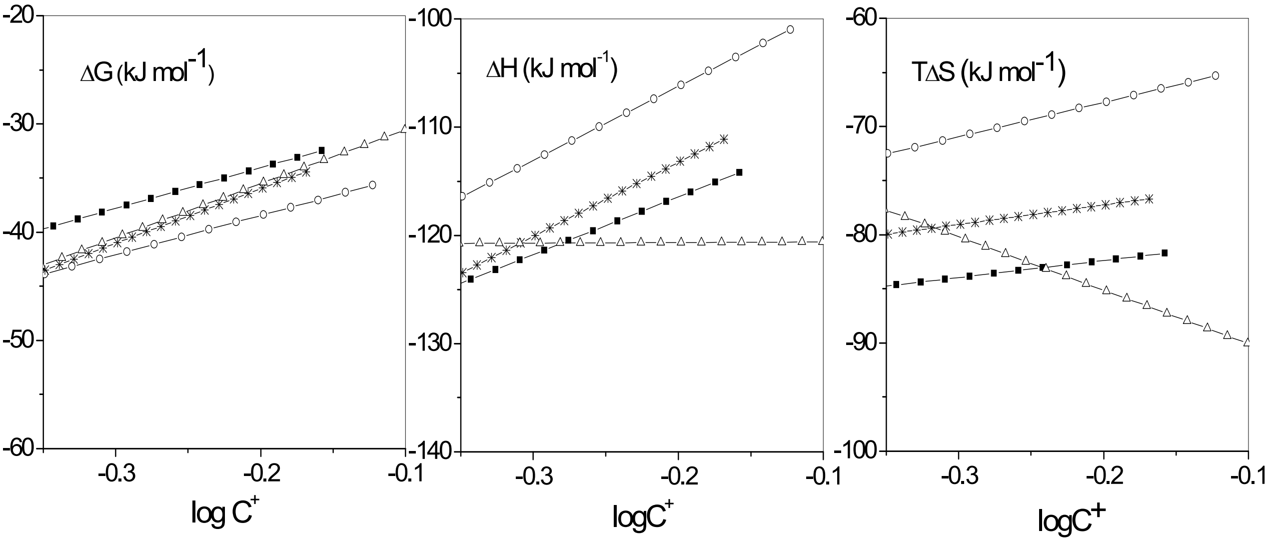

Figure 3.21. Influence of the cation concentration on the measured thermodynamic parameters of the Max p21/E-box binding reaction at 25 °C. All parameters change linearly with log[C^+]. For clarity only the fitted lines are presented. Circles, Max p21^{N19A}. Triangles, Max_{short}^{SS}, Squares: Max p21 wild type. Asterisks, Max p21VL. The data are scaled identically, so that the slopes of the lines in the panels can be compared directly.

Altogether, the data strongly suggest that the Asn 19 side chain is not involved in electrostatic contacts with the DNA backbone, while additional protein-phosphate backbone interactions are responsible for the higher-affinity binding of the variants with stabilized b-HLH-LZ domain. We are reluctant to embark on lengthy speculations as to which side chains may be involved. In principle, there are many highly-conserved basic residues in close proximity to the DNA backbone, involving the presumably quite unrestrained stretch Lys 13-Arg 14-Arg 15 preceding the “canonical” portion of the basic region (His 18 to Arg 26), Lys 24 in the middle of the basic region, Lys 30 from helix H1, as well as Lys 47 from the loop region. Indeed, some of these interactions were described in, or can be surmised from the crystal structures of the c-Myc-Max-E-box and $(Max)_2$-E-box complexes, both containing only the b-HLH-LZ protein domains (*52*).

The feasibility of these contacts is less obvious in the only available structure of the Max p21-E-box complex. It is perhaps most naturally to assume that a more stable and possibly less fluctuating leucine zipper alters the structural or dynamic properties, or both, of the adjacent HLH domain. It has been argued that stabilization of structural scaffolds or stabilization of local conformations could improve the DNA-binding affinity of transcription factors (*78*). Possibly, conformational rearrangement of the loop region and slight positional reorientation of helices H1 and H2 facilitate new protein-DNA contacts. In fact, the NMR structure of the free, disulfide linked Max b-HLH-LZ domain bearing the N78V/H81L double mutation indicates a loop conformation facilitating direct contact of Lys 47 and Ser 49 to DNA phosphates (*50*). Lys 47 points away from the duplex in $(\mathrm{Max})_2$-E-box X-ray structure. As another, very likely candidate we suggest Lys 30 of helix H1, which forms very well defined charge-charge contacts with DNA phosphates (distance < 3.5 Ǻ) in both monomers of the c-Myc-Max-E-box complex.

The calorimetric method gives the unique opportunity to gain insights into the thermodynamic signature of the salt effect. The effect of salt on protein-DNA interactions is usually treated in the framework of the counter-cation/limited law model (CC/LL; (*70*)]. Weaker binding at higher salt concentration is thought to originate from the less favourable entropy of expelling DNA-associated cations into the bulk. At high salt concentration there is less entropy gain from cation displacement. According to CC/LL, the enthalpic effect is believed to be close to zero. Figure 3.21 illustrates how ΔG_A, ΔH_A and $T\Delta S_A$ change in response to the increasing salt concentration. Only $\mathrm{Max}_{\mathrm{short}}^{\mathrm{SS}}$ follows the prediction of the CC/LL model: ΔS_A decreases at higher salt concentration, ΔH_A being roughly insensitive. In all other variants the decrease in affinity at high salt is dominated by a decrease of ΔH_A. In fact, $T\Delta S_A$ gets more favourable as the cation concentration increases, compensating partly the enthalpic loss (compare the slopes of the functions in Figure 3.21). Enthalpic weakening of protein-DNA complexes has been observed before (*79*). Theoretical models based on evaluation of the electrostatic salvation energy of polyelectrolytes by the Poisson-Boltzmann equation predict a sizeable enthalpic loss from salt expulsion (*80, 81*). However, our data showing a dominant role of ΔH_A, in conjunction with smaller and reversed entropic effect are in sharp contrast with the available information. What could cause the strong enthalpic destabilization? Indeed, binding includes a large conformational change of the basic region. However, similar conformational changes in the GCN4 transcription factor do not cause deviation from the CC/LL model (*75*). The overall structure (as judged by CD spectroscopy) and the thermodynamic stability (as judged by thermal melting) of the proteins are not detectably influenced by salt. ΔH_A linearly extrapolated from high-salt conditions to the standard working buffer conditions are in excellent agreement with ΔH_A directly measured by ITC. We believe that the likely reasons are osmotic effects at the relatively high salt concentration used in this study. Binding involves water molecules trapped at the protein-DNA interface, as seen by x-ray crystallography (*47, 48, 52*). The changes in the bulk salt concentration might lead to redistribution of waters. It is known that water molecules serving as bridges between protein and ligand groups might contribute to the enthalpy of binding (*82*). Experiments with "dedicated" osmotic agents (glycerol, sucrose, betaine) are necessary to probe for the influence of a general osmotic effect. Specific anion binding could also be contributor (*83, 84*). The clearly different behaviour of $\mathrm{Max}_{\mathrm{short}}^{\mathrm{SS}}$ is puzzling and might indicate the role of the frying ends of the leucine zipper sub-domain to the observed energetic effect. In any case, the Max p21/DNA complex provides a promising system to study the effect of salt on the energetics of macromolecular recognition.

3.2.5 Mutational analysis of the Max p21/E-box complex

Based on structural information six residues from the highly conserved basic region of the Max p21 monomer contact the Cyt(1)-Ade(2)-Cyt(3)-Gua(4)-Thy(5)-Gua(6) E-box core sequence (*47*). Although more contacts could be inferred from the X-ray structures of $(\mathrm{Max})_2$-E-box, Myc-Max-E-box and Mad-Max-E-box complexes formed by the corresponding b-HLH-LZ core domains, we analyze the contacts as they are defined in the only available complex of the full length Max p21. Residues His 18, Glu 22 and Arg 26 are involved in specific hydrogen bonds to DNA bases. Mutation of Glu 22 abolishes discrimination between E-box and non-specific DNA (*47*). The contact formed by Arg 26 to the central guanine crucially governs the specificity of b-HLH-LZ proteins for class B E-box elements (*47*). Arg 26 is anchored also to the DNA backbone since both NH2 and Nε atoms are positioned very close to phosphate oxygen atoms. Non-specific interactions with the phosphate backbone are formed by Asn 19, Arg 23 and Arg 25. Outside the basic region, Arg 50 which is located at the start of helix 2, makes additionally a main-chain and a side-chain contacts with backbone phosphates. A schematic representation of the interactions is shown in Figure 1.6. To probe for the energetic role of particular contacts we replaced the corresponding side chains by alanine and determined the energetic signature of the mutant complexes by ITC. Examples of ITC binding experiments are shown in Figure 3.22.

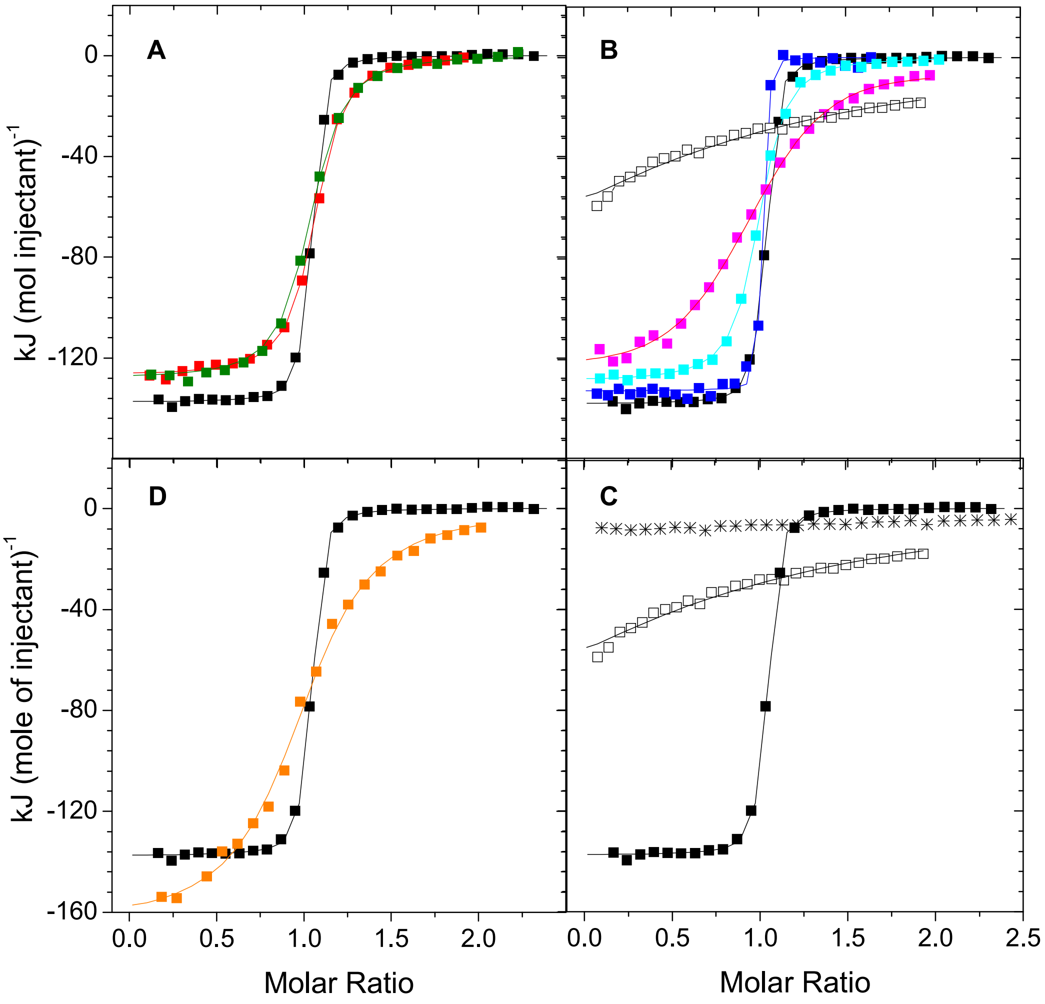

Figure 3.22. Binding isotherms obtained by ITC at 25 °C for binding of Max p21 X-to-Ala mutants to E-box DNA. In all panels data collected with wild type Max p21 are presented for reference (black symbols and black lines). A, Mutations removing base-specific hydrogen bonds. Red, Max p21^{H18A}. Green, Max p21^{E22A}. B, Mutations removing protein-DNA backbone contacts from the basic region. Blue, Max p21^{N19A}. Cyan, Max p21^{R23A}. Magenta, Max p21^{R25A}. C, Data collected with Max p21^{R50A} are shown in brown. Open symbols in panels C and D represent titrations with Max p21^{R26A}. Titration data with wild type Max p21 and unspecific DNA are shown with asterisks in panel D. Symbols are the integrated heats at each titration step. The solid lines are best non-linear regression fits according to 1:1 binding model.

The results with Max p21 are shown as a reference. As discussed in section 3.2.2 these isotherms (as well as the isotherms obtained with Max p21^{N19A}) are too steep and can not be used for calculation of the binding constant. The measured thermodynamic parameters at 25 °C are listed in Table 3.8. Except for wild-type Max p21 and Max p21^{N19A} these parameters are directly determined by ITC. However, for several mutants binding at lower temperatures becomes too tight (Figure 3.23).

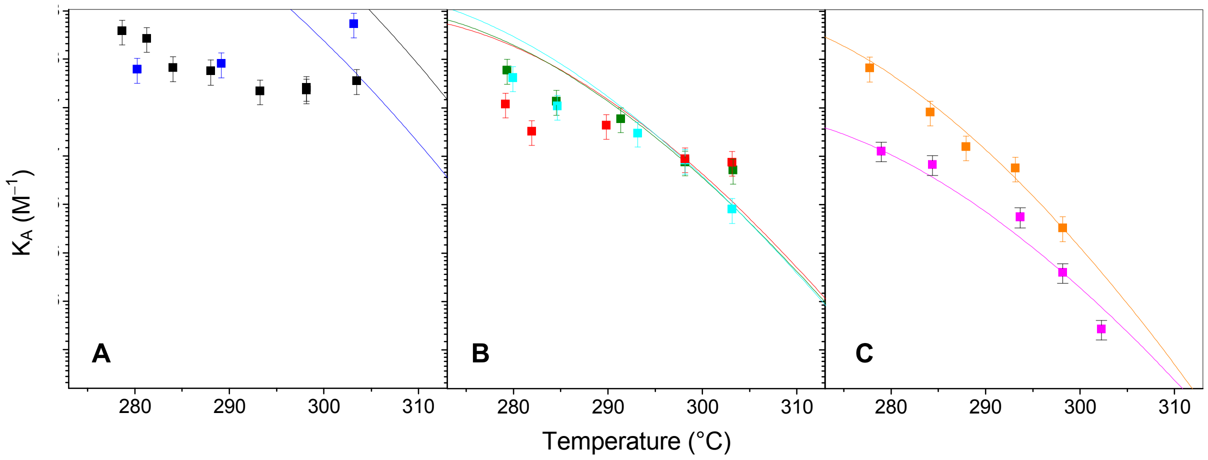

Figure 3.23. Temperature dependence of the apparent and genuine equilibrium constants describing binding of Max p21 variants to E-box DNA. K_A is plotted on a natural logarithmic scale identical for each panel to facilitate comparisons. Apparent K_A obtained from non-linear regression analysis of individual ITC isotherms are shown with symbols (± the mean standard deviation in the set). The lines represent the calculated K_A = f(T) function according to equation 3.8 using the parameters listed in Table 3.8. as the reference parameters. A, Max p21 (blue) and Max p21^{N19A} (black). B, Max p21^{H18A} (red), Max p21^{E22A} (green) and Max p21^{R23A} (cyan). C, Max p21^{R25A} (magenta) and Max p21^{R50A} (brown).The affinity decreases from A to C. Note that for strong binding, the apparent K_A underestimates the genuine affinity. The correspondence between the calculated function and the experimental data improves as the affinity decreases.

The thermodynamic profiles of these mutant complexes were constructed with help of equation 3.8. The reliability of the parameters listed in Table 3.8 as reference values in equation 3.8 is demonstrated with Max p21^{R25A} and Max p21^{R25A} which are the weakest binders, so that direct determination of K_A between 5 °C and 25 °C was possible (Figure 3.23).

Figure 3.24 illustrates the changes in the energetic profile of the Max p21/E-box complex caused by the mutations. The bars represent ΔΔG, ΔΔH and TΔΔS values ($\Delta\Delta X = \Delta X^{mutant} - \Delta X^{wild\ type}$) calculated for binding of the Max p21 dimer to E-box duplex, that is, the indicated energetic differences are the combined effect of two simultaneous mutations. We do not know whether the structural symmetry seen in the crystal structure translates into energetic equality of the contacts realized in the two half sites. Removal of the base-specific hydrogen bond of Arg 26 by replacing the central E-box guanine by cytosine in one or both E-box half sites has demonstrated essentially no cooperativity (*47*). This result should not be taken as ultimately proving the independence of the half sites since all other protein-DNA contacts have been left intact in the cited study.

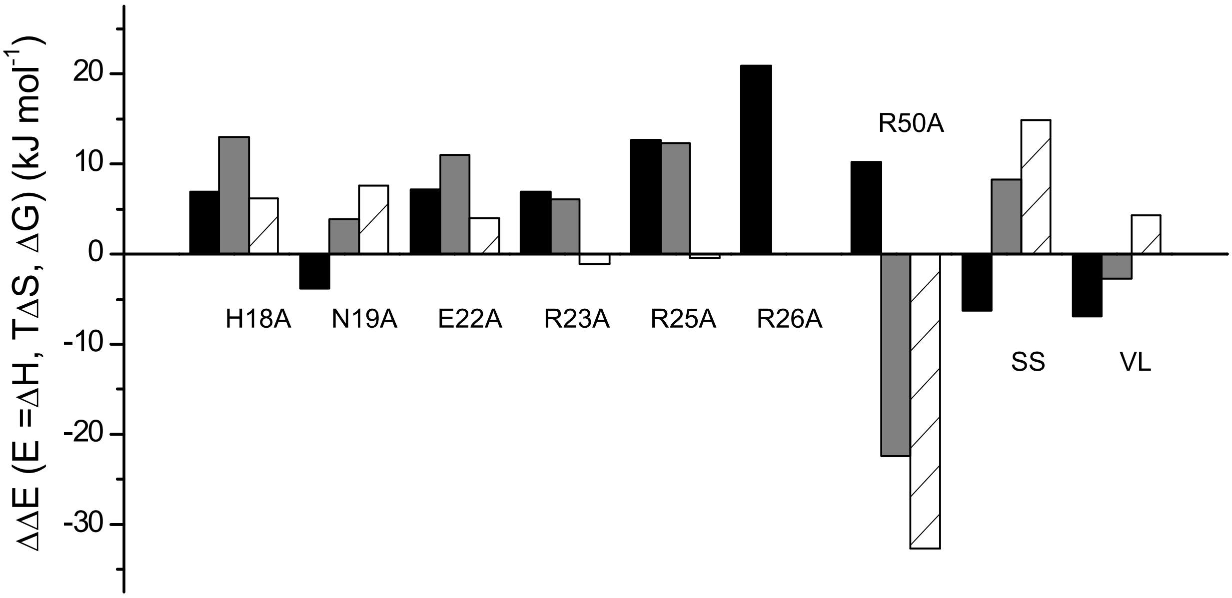

Figure 3.24. Changes in the energetic profile of the Max p21/E-box complex upon alanine substitution or LZ stabilization. The bars represent the changes in ΔG_A (black), ΔH_A (grey) and $T\Delta S_A$ (hatched) calculated as $\Delta\Delta E = \Delta E^{mutant} - \Delta E^{wild\ type}$ (numerical ΔE values from Table 3.8). $\Delta\Delta G > 0$ indicates that the mutation is destabilizing. Mutations are destabilizing the complex enthalpically if $\Delta\Delta H > 0$. Entropically favourable mutations are manifested by $T\Delta\Delta S > 0$. All experiments were performed at 25 °C in the standard buffer.

Small distance differences between interacting groups might result in serious energetic differences. However, we are not aware of any study documenting cooperativity between half sites in DNA recognition by (structurally) symmetrical homodimer proteins. Three mutations, H18A, E22A and R23A exhibit modest decrease of affinity (~sixteen time increase in K_D). The effect of removal of Arg 25 and Arg 50 is more pronounced (a K_D increase of 160 and 60 times, respectively. Two mutations stand out. The wild type Asn 19 side chain appears to destabilize complex relative to alanine. Removal of the Arg 26 side chain beyond the C_β atom is extremely destabilizing. We can not give a very precise estimate for the decreased affinity of the Max p21^{R26A} protein. ITC experiments were not possible at concentrations sufficiently high as to allow precise measurement of the binding constant at 25 °C. The observed heat effect of binding was small as well. These experiments indicated K_D ~ 50 μM as the lower bound. Since the pronounced reduction of heat release relative to wild type binding introduced ambiguity in the ITC results, another estimate of the Max p21^{R26A} affinity was obtained by CD spectroscopy. Recently, it was convincingly demonstrated that the decrease in MRE_{222} upon titration of Max p21 with increasing amounts of DNA directly reflects the population of the protein/DNA complex (*85*). Mixing of equimolar (50 μM) Max p21^{R26A} and E-box duplex results in a decrease in MRE$_{222}$ of 400 mdeg cm dmol^{-1}. This is ~15 % of the average MRE$_{222}$ decrease observed at full saturation (3000 mdeg cm dmol^{-1}). With this information, K_D ~200 µM was calculated. Whatever the exact number might be, the R26A mutation destabilizes the complex by a factor of at least 10^4 (K_D).

Is it possible to rationalize the energetic effect of the mutations? Visual inspection of the protein/DNA co-crystal structure provides clues for some sites. Removal of the four base-specific and buried hydrogen bonds formed by Glu 22 to N4 of Cyt(1) and N6 of Ade(2) (two hydrogen bonds per half site) is energetically not very costly: Each bond contributes 1.7 kJ mol^{-1} on average to ΔG_A. Interestingly, if the bonds are broken by mutating out the hydrogen donors, the average contribution per hydrogen bond is more than twice larger, 3.7 kJ mol^{-1} (*85*). It is likely that the carboxyl group of Glu 22 experiences large unfavourable dehydration, which is uncompensated in the absence of hydrogen bond partners. The problem appears relieved in the Max p21^{E22A} mutant lacking the side chain carboxylate. The dramatic destabilization caused by the Arg 26 replacement is surprising on a first glance, since the contribution of each base-specific hydrogen bond between NH1 of Arg 26 and N7 of Gua(4) has been recently estimated as 2-3 kJ mol^{-1} (*53*). It should be noted, however, that the Arg 26 side chain is engaged additionally in two backbone contacts, which are also lost (broken) in the R26A mutant studied here. Apart from the energetic contribution of each backbone contact per se (which can be substantial as seen in Figure 3.24) such interactions might help to fix the arginine side chain in a position for optimal hydrogen bonding and certainly help to overcome the entropic lost from freezing of the side chain. In any case, Arg 26 represents a crucial “hot spot” at the binding interface. Overall, there is a large variation in the destabilizing effect caused by the removal of interactions. On average, backbone contacts contribute more to stabilization of the complex than base-specific hydrogen bonds. This is in line with previous results from alanine scanning showing that substitutions of residues inferred from the co-crystal structure to make contacts with the sugar–phosphate backbone may produce mutant proteins more severely defective in DNA binding than substitution of residues thought to make base-specific contacts (e.g. (*86*)). The energetic content of such “unspecific” interactions might also differ considerably. A clear example is presented by Arg 23 and Arg 25, the effect of the latter being twice larger. In fact, Arg 23 is oriented much more favourably toward the phosphate oxygen atoms than Arg 25. However, it is also significantly more exposed to the solvent than Arg 25. Moreover, we note that the Arg 25 $N\varepsilon$ atom faces the carboxylic group of Glu 22 (it is < 3.5 Å away from both carboxylate oxygens). This observation suggests that the Arg 25 side chain may help in optimizing the geometry of the base-specific contacts of Glu 22. If so, it exerts an “indirect” additional stabilization of the complex, which is lost upon mutation to alanine. Particularly intriguing is the role of Asn 19. The residue is highly conserved within the Myc-Max-Mad network of transcriptional factors and it closely approaches the DNA backbone according to structural data, yet the removal of this contact by mutating out the asparagine amide group increases the affinity for DNA. According to the original annotation of the available coordinate files the contact to the phosphate oxygen is made by either OD1 or ND2 atoms of Asn 19 (distance < 3 Å). Since the assignment of the electron density to OD1 or ND2 atoms in not very highly resolved structures is ambiguous, we have attempted a simple visual evaluation of the possibility that the potentially unfavourable Asn OD1-to-O1P contact can be relieved by rotating the C_β-C_γ bond to position ND2 close to O1P. No clear conclusions could be drawn. It appears that the environment of the Asn 19 amide group is electrostatically intensive due to the close proximity of groups bearing partial positive charge (ND1 of His 18) or partial negative charge (the carbonyl oxygen of Arg 15) and bound water molecule(s). The problem requires computational investigation.

Concerning the energetic partitioning of the changes in affinity, we can not provide a detailed structural explanation of the observed enthalpy-entropy balance of particular mutations. First, the experimental error is sizeable in comparison to magnitude of the measured $\Delta\Delta H_A$ and $\Delta\Delta S_A$. Second and more important, binding is tightly coupled to (partial) coil-to-helix transition of the Max p21 basic region and to bending of the E-box duplex. In principle, the energetic signature of these processes could be affected by mutation. Third, replacement by larger side chains having polar groups by small and non-polar alanine could have caused redistribution of water molecules at and near the binding interface. With these considerations in mind, in the following we briefly discuss some general trends and possible sources of the energetic signature of some mutations.

We first consider mutations in the basic region, excluding the R26A mutation, for which $\Delta\Delta H_A$ and $\Delta\Delta S_A$ could not me measured. The R50A replacement exhibits a completely different energetic signature and should be discussed separately. It is clear from Figure 3.24 that all mutations in the basic region are linked to loss of enthalpic interactions, including the stabilizing N19A replacement. For side chains stabilizing the protein-DNA complex, the magnitude of enthalpic destabilization upon mutation dominates over the entropic effect, but there is no correlation between the total destabilization ($\Delta\Delta G$) and its enthalpic ($\Delta\Delta H$) and entropic ($\Delta\Delta S$) components, as noted in diverse other macromolecular systems. Removal of a backbone contact could be enthalpically as costly as the removal of two hydrogen bonds (R25A *versus* H18A and E22A). The enthalpic contribution from formally identical backbone contacts could also be very different (R23A *versus* R25A). In fact, the enthalpy term completely determines the magnitude of stabilization provided by the latter two arginine side chains.

The changes in heat capacity upon mutation are quite small (5 to 13 % of $\Delta C_{p,A}$), within the experimental uncertainty (Table 3.8). It should be considered, however, that the total heat capacity decrement is largely dominated by the refolding of the basic region. Since single alanine mutation are not expected to alter the coil-to-helix transition of the basic region significantly, and the “rigid-body” heat capacity change is rather small (–1 kJ mol^{-1} K^{-1} or smaller), the measured $\Delta\Delta C_{p,A}$ values become sizeable. There is no correlation between the experimental $\Delta\Delta C_{p,A}$ and the amount of polar, non-polar, and total buried surface change caused by mutation. Interestingly, $\Delta\Delta C_{p,A}$ strongly correlates with the free energy effect of the mutations ($\Delta\Delta G_A$). The stabilizing N19A replacement is linked to the largest negative heat capacity increment (–0.4 kJ mol^{-1} K^{-1}), while the most destabilizing mutation (R25A) is accompanied by the largest positive $\Delta\Delta C_{p,A}$ (0.5 kJ mol^{-1} K^{-1}). Overall, $\Delta\Delta C_{p,A}$ and $\Delta\Delta G_A$ are correlated by $R^2 = 0.96$. This observation possibly indicates a complicated balance of forces and mechanisms leading to change in affinity, since neither differential parameter (including $\Delta\Delta H_A$ and $\Delta\Delta S_A$) correlates with the change of surface burial upon mutation, and there are no other cross-correlations between $\Delta\Delta X$ parameters.

For all sites entropic factors disfavour the wild type side chains or are negligible. The entropy gain can not be explained with the gain of side chain entropy when longer side chains are replaced by alanine, since the wild type arginine side chains appear not penalized entropically. We note the general trend that the favourable entropic effect decreases the closer the site of mutation is to the point where the basic region emerges from helix H1. Sauvé et al. have demonstrated by NMR the existence of a persistent helical turn spanning residues Arg 25 to Ile 29 in the absence of DNA, even at 35 °C (*85*). Furthermore, the Arg 25 and Arg 26 side chains adopt an extended conformation, which is virtually identical to the conformation observed in the Max p21-E-box complex. It is therefore naturally to assume that substituting either residue by alanine (having higher, yet very similar helical propensity as arginine) would not interfere with both the fractional population of the helical turn and the conformation of the adjacent arginine side chain. Differently, His 18, Asn 19, Glu 22 and Arg 23 obtain a stable α-helical conformation only in the context of the protein-DNA complex, but may exist in transient α-helical states also in the free protein. As far as stabilization of such states spanning the basic region in the free protein would decrease the entropic penalty for finding the proper conformation facilitating formation of intermolecular contacts, alanine mutation in the corresponding sites is expected to stabilize the complex entropically. Figure 3.25 illustrates the very strong correlation between the increase in α-helical propensity upon X-to-Ala mutation and the entropic benefit the mutation exhibits. However favourable alanine substitutions might be in terms of entropy, the *total* energetic effect of mutation is greatly modulated by the enthalpic factors (i.e. formation/breakage of bonds and the concomitant hydration changes). This point is nicely illustrated by comparing the energetic signature of H18A and N19A mutations.

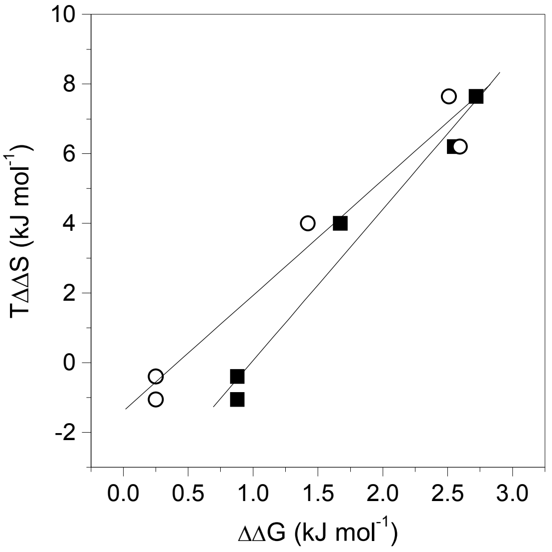

Figure 3.25. Correlation between the measured change in entropy upon X-to-Ala substitutions of side chains in the basic region of Max p21and the α-helical propensity of the replaced side chains. TΔΔS is plotted against ΔΔG, which ranks the amino acid side chains according to their preference to adopt α-helical conformation (low ΔΔG indicates high α-helix propensity, relative to alanine, for which ΔΔG = 0). The helix propensity scales derived entirely from experimental data with peptides and proteins are shown. Filled symbols, scale by Pace & Scholtz (R^2 = 0.97:Pace et al., (1998), Biophys. J. 75, 422-427). Open symbols, scale AGADIR (R^2 = 0.96; Lacroix et al., (1998), J.Mol.Biol. 284, 173-191)

The energetic effect of the R50A mutation is balanced in a radically different way. First, both the enthalpy and the entropy changes are much larger than the changes observed for all other sites. Second, the replacement is favoured by a large enthalpic effect and is disfavoured by an even larger entropic effect. Since the mutation eliminates only one of the two contacts of Arg 50 to the phosphate backbone (the second one being formed by the peptide amide) this means that the interactions formed by the Arg 50 side chain destabilize the wild type complex enthalpically, yet strongly promote DNA binding entropically. Both the magnitude and the sign of the observed energetic changes are very unlikely caused by elimination of a single contact to the phosphate backbone. Rather, we envisage contributions from structural rearrangements of the protein in response to the mutation. The methylene groups of the Arg 50 side chain of pack against Ile 29 and Phe 33 and complement the HLH hydrophobic core. These interactions are present both in the X-ray structure of the Max p21-E-box complex and in the NMR ensemble of the free Max protein but are eliminated in the Max p21^{R50A} mutant. Two scenarios are possible. (i) Removal of the hydrophobic moiety of Arg 50 side chain by alanine mutation causes loosening of the packing of the HLH hydrophobic core. In the Max p21^{R50A}-DNA complex, the contact made by the peptide amide to the phosphate backbone promotes consolidation of the packing interactions in the region. (ii) Alternatively, strong contacts to DNA made simultaneously by the Arg 50 guanidino and amide groups might lead to a “conformational strain” and suboptimal packing of the side chain methylene groups to Ile 29 and Phe 33. If the second contact to DNA is removed by mutation, no structural changes are imposed in the region. It is not possible to distinguish between the two scenarios, yet both are fully compatible with the observed energetic signature of the R50A mutation. Loosening (suboptimal packing) of hydrophobic interactions in the HLH core is linked to disruption of enthalpically rich contacts, but a less tightly packed, and possibly more flexible region will be entropically favoured. Also, although the difference is small in magnitude, $\Delta C_{p,A}^{R50A}$ is more negative than $\Delta C_{p,A}^{WT}$, as expected for a process linked to exposure of hydrophobic surface. Not only for this site, a rigorous analysis of the energetic partitioning of the observed changes in affinity arising from structure perturbation is unfortunately hampered by the lack of high-resolution structural information on all relevant states comprising the thermodynamic system.

Table 3.8. Thermodynamic parameters characterizing Max p21 and Max p21 variants binding to E-box DNA measured by ITC.[a]

protein	$K_A \times 10^{-7}$ (M^{-1})	ΔG_A (kJ mol^{-1})	ΔH_A (kJ mol^{-1})	$T\Delta S_A$ (kJ mol^{-1})	ΔS_A (J $K^{-1} mol^{-1}$)	$\Delta C_{p,A}$ (kJ $K^{-1} mol^{-1}$)
Max p21	38.00 69.00[b]	–48.9	–137	–88	–295	–3.8
Max p21VL	610.00	–55.8	–140	–84	–282	n.d.
Max_{short}^{SS}	480.00 310.00[b]	–55.2	–129	–74	–248	–3.8
Max $p21^{H18A}$	2.30	–42.0	–124	–82	–275	–3.8
Max $p21^{N19A}$	180.00	–52.8	–133	–80	–268	–4.2
Max $p21^{E22A}$	2.15	–41.8	–126	–84	–282	–3.6
Max $p21^{R23A}$	2.23	–41.9	–130	–88	–295	–3.6
Max $p21^{R25A}$	0.22	–36.2	–125	–89	–300	–3.3
Max $p21^{R26A}$	$< 10^{-3}$	> -25.0	n.d.	n.d.	n.d.	n.d.
Max $p21^{R50A}$	0.55	–38.5	–159	–119	400	–4.1

[a] All experiments were performed at 25 °C in 100 mM Na-phosphate, 100 mM KCl, pH 6.8, unless otherwise indicated. The standard error of K_A, δK_A was maximum 60%. Thus, the maximum error of ΔG_A is $\delta\Delta G_A = RT$□□$(\delta K_A/K_A) = 1.5$ kJ mol^{-1}. The maximum error of ΔH_A, $\delta\Delta H_A$, was on the order of 6 kJ mol^{-1}. The error of $T\Delta S_A$ is $\delta T\Delta SA = \sqrt{(\delta\Delta H_A)^2 + (\delta\Delta G_A)^2}$ = 6.5 kJ mol^{-1}. The error of $\Delta C_{p,A}$ was estimated as ±0.2 kJ K^{-1} mol^{-1} by jack-knife tests. [b] From DSC data ac to equations 3.7-3.9.

3.2.6 Does monomeric Max p21 recognize the E-box target?

According to structural data, dimeric transcription factors bind to a DNA site with dyad symmetry, each monomer recognizing one half-site. Since most of them capable to dimerize in the absence of DNA, it has been thought that dimerization is obligate in order to recognize the DNA target site. In the last decade this view was challenged on the basis of experimental data suggesting that the observed rapid rates of DNA binding could not be accounted for if formation of a dimeric bZIP peptide had to precede DNA binding (*62*). Thus, a monomeric binding pathway was proposed, according to which monomers first bind to a half-site and dimerize while bound to the duplex. Apart from the potential increase of the binding rate, the biological advantages of the monomer binding pathway are obvious. Even not very stable dimers could assemble on the DNA and trigger the biological effect at low concentration. Furthermore, specificity may be increased, trapping of the transcription factor at non-specific DNA sequences might be avoided, different DNA sites might be targeted by heterodimeric proteins. The existence, and the dominant role of the monomeric pathway has been documented for dimeric transcription factors belonging to different classes, including the LexA and Arc repressors, b-ZIP, b-HLH and b-HLH-LZ proteins (*60-65*)

Indeed, kinetic experiments have confirmed that the Max b-HLH-LZ core domain binds to E-box DNA as a monomer (60, 64). Using a thermodynamic cycle to derive K_D, Hu et al. recently suggested that monomeric Max b-HLH-LZ binds to DNA with very low nanomolar affinity (8 nM; (*60*)), comparable with the affinity of the dimeric protein. In contrast, the stability of an initial Max b-HLH-LZ monomer/DNA complex has been previously estimated as 1 µM (*64*). According to data collected in our laboratory (61), the monomeric b-region of the GCN4 protein binds to the AP-1 target with K_D in the range 0.4 µM (derived from kinetics) to 2 µM (derived by ITC). Relatively weak monomer DNA-binding affinity (>10 µM) was also suggested for the monomer of the closely related activating transcription factor ATF-2 (*64*).

Intrigued by the suggested very high affinity of the Max b-HLH-LZ monomer for E-box DNA, we designed, expressed and purified a largely monomeric version of the Max p21 isoform. The deeply buried Leu 36 participating in the hydrophobic core of the dimeric HLH domain was replaced by aspartic acid, intending disruption of the dimerization interface. Since the oligomerization behaviour of this mutant could not be established unequivocally, a second mutation was introduced by replacing Met 64 at the beginning of the leucine zipper. Figure 3.26 presents data on the biophysical characterization of Max p21$^{L36D/M64P}$ protein. AUC experiments show that the dimer-monomer dissociation constant is on the order of 100 µM, meaning that the protein is predominantly monomeric at the very low micromolar concentrations utilized in the

ITC experiment. Judging by CD data, while not completely in a random coil conformation, Max p21$^{L36D/M64P}$ exhibits a drastic reduction in helical content.

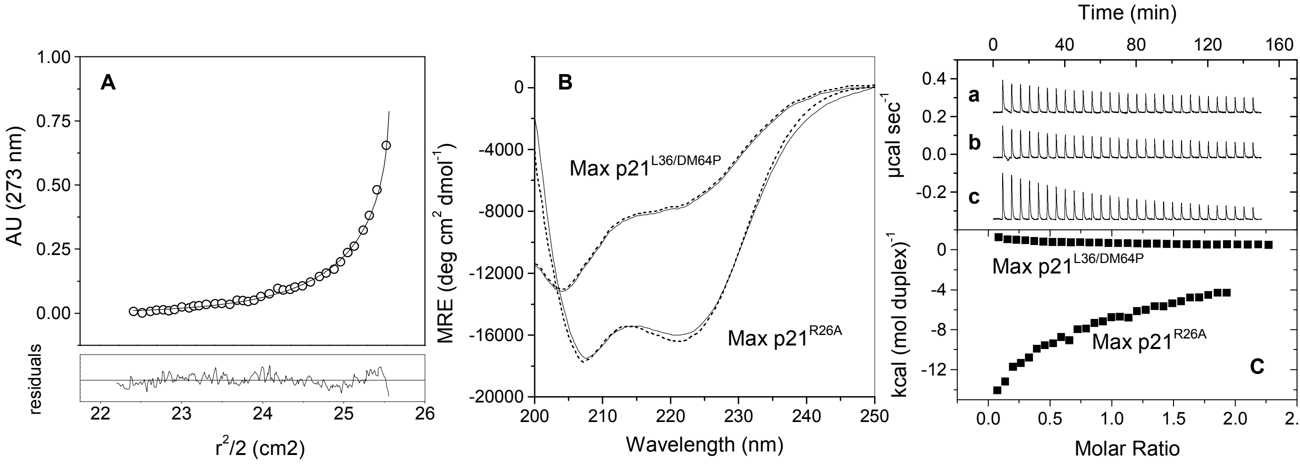

Figure 3.26. Biophysical characterization of the Max p21$^{L36/DM64P}$ variant. (A) Analysis of the ologomerization state at 25 °C by sedimentation equilibrium analysis. Upper panel, UV absorbance gradient as function of the radial position (circles) and the fit according to a single species model (line). Lower panel, Residuals showing the difference between the experimental data and the theoretical model. The concentration was 500 µM (monomer equivalents). The data can not be described with a single species model. Rather, (B) Far UV CD spectrum. Concentrations in all experiments were 50 µM protein dimer, E-box duplex and 1:1 protein-DNA complex. Temperature was 20 °C. Continuous lines, free protein. Dotted lines, protein-DNA complex. (C) ITC binding experiments. Upper panel, Raw calorimetric signal observed in titration of Max p21$^{L36/DM64P}$ (480 µM monomer equivalents) into plain buffer at 7 °C (a), into 24 µM DNA duplex at 7 °C (b), and Max p21$^{L36/DM64P}$ (520 µM monomer equivalents) into into 24 µM DNA duplex at 25 °C (c). The thermograms are shifted on the y-axis for clarity. Lower panel, the integrated heats in the experiment at 25 °C. For comparison, the results from titration of Max p21^{R26A} (120 µM monomer equivalents) into 12 µM DNA duplex at 25 °C are shown.

We tested the binding properties of the monomeric Max p21 by ITC. As illustrated in panel C of Figure 11, weak endothermic signals were observed both at 7 and 25 °C. The resulting very shallow isotherms could not be deconvoluted with a stoichiometric binding model. Moreover, blank titrations of protein into buffer produced essentially identical thermograms, indicating that the slight decrease of the integrated heats at higher protein:DNA ratios is not caused by saturation of binding sites. Rather, it reflects the shift in the dimer-monomer equilibrium as the highly concentrated solution in the injection syringe becomes diluted in the calorimetric cell. Since, in principle, the total apparent heat effect of binding might be below the limit of detection, we tested whether the helical content of Max p21$^{L36D/M64P}$ increases in the presence of DNA, as to indicate refolding of the basic region upon binding. No change was observed. Panels B and C in Figure 3.26 show the results of ITC and CD experiments with the Max p21^{R26A} mutant, for which a K_D > 50 and close to 200 µM was estimated (see above). It appears that Max p21$^{L36D/M64P}$ binds even weaker, if at all. In summary, the described experiments let us conclude that the basic region in the context of full-length unfolded Max p21 monomer is incapable to promote significant recruitment of Max monomers to the E-box site in and below the low micromolar concentration range.

We are well aware that the presented results do not invalidate any previous data corroborating the monomer binding pathway, a valuable evolutionary solution. By design, the monomeric pathway leading to the final complex is disallowed, sinceMax p21$^{L36D/M64P}$ is not capable to form dimers at all, and can not assemble to a dimer even when the two monomers are bound to the duplex. Although not very likely, it is also possible that the mutations disrupt some structures existing in the "native" unfolded state and promoting monomer binding. Nevertheless, the apparent lack of Max p21$^{L36D/M64P}$ binding might point to the role of protein segments outside the boundary of the b-HLH-LZ domain. For example, the full-length unfolded state might be more compact than the much shorter unfolded b-HLH-LZ, thus preventing or slowing down formation of b-region-to-DNA contacts. Also the rate of diffusion of the unfolded full-length protein is expected to be lower. Further investigations are required to explain the discrepant results on the stability of the Max monomer/E-box complex. There is no kinetic information about none of the elementary steps along the Max p21 monomer and Max p21 dimer binding pathways. Hence, it might turn out that the monomeric pathway is still relevant in vivo, irrespective of weak monomer binding to the E-box target.

3.2.7 Conclusions and outlook

We have characterized the energetic basis of site-specific E-box recognition by the full length gene product of the Max transcription factor p21 isoform. The N- and C-terminal segments flanking the core b-HLH-LZ domain and accounting for half of the protein are unstructured, do not participate in packing interactions, and are not involved in DNA binding. Their role in promoting the biological effects of Max is still poorly understood. Max p21 binds the E-box target with affinity in the low nanomolar range at 37 °C. The binding reaction is driven by net favorable enthalpy changes and is opposed by the net entropic changes. The refolding of the unstructured basic region of the free protein to the binding-competent α-helical state significantly contributes to the overall thermodynamic parameters. Stabilization of the LZ subdomain increases the affinity of Max p21 to DNA. The results suggest that thermodynamic coupling between HLH and LZ leads to subtle structural changes, as the consequence of which two Max residues form additional electrostatic contacts with the DNA phosphates, the likely candidates being Lys 30 from helix H1 and Lys 47 from the loop.

We probed the energetic importance of evolutionary conserved residues that form well defined contacts to the E-box target DNA bases or backbone phosphates according to X-ray data. Five side chains contribute modestly to binding affinity (15-160× decrease in K_D upon alanine mutation). One notable exception is Arg 26, which confers substantial stabilization. Although in close contact to the phosphate backbone, the Asn 19 side chain appears energetically unimportant. Altogether, the presented data point to the central role of the helical turn encompassing the C-terminus of the basic region and the start of helix H1. The pre-formed bivalent helical scaffold possibly anchors the protein to DNA at little expense of conformational entropy by positioning the energetically most important Arg 25 and Arg 26 to form three backbone contacts and one base-specific, deeply buried hydrogen bond. Association may be fast and dissociation may be slow since the clustered positive charges of Arg 25, Arg 26 and Arg 50 provide a steering force toward DNA. Furthermore, once stabilized, this conformation might serve as the nucleation site for propagation of α-helix toward the N-terminus of the basic region. Analysis of the thermodynamic signature of alanine mutants suggests that there are pronounced, context-dependent differences in the energetic content of formally identical protein-DNA contacts (for instance arginine-backbone phosphate bonds). Finally, the results with the unfolded Max p21 monomer are at odds with previous data corroborating the relevance of the monomer DNA-binding pathway, and prompt for further investigations of the relevance of the monomer binding pathway in site-specific E-box recognition by b-HLH-LZ transcription factors.

Since the degree of sequence conservation in the basic and loop regions of Myc/Mad/Max family members is not absolute, future results from structural and biophysical experiments, and molecular mechanics approaches will help to reveal further details on the mechanism and energetics of E-box recognition by Max p21.

The following lines of research are planed (or are already followed) in this laboratory. Molecular dynamics simulations identify significantly populated protein-DNA contacts, which are not unequivocally defined in the existing X-ray structures of the Max b-HLH-LZ/E-box complexes. Several arginine and lysine side chains have been already mutated to alanine, and the stability of the mutant complexes is being investigated. DNA binding experiments with the Myc/Max and Mad/Max heterodimers will provide information on whether the in vivo relevant competition for the target E-box is based on discriminating DNA binding affinity, or else is an indirect consequence of the stability of the dimers, i.e. regulation is promoted mainly by expression level control. Protein and DNA variants bearing fluorescence reporter groups will be designed and prepared as tools to obtain information on the rates of DNA binding/unbinding and of protein dimerization/dissociation, and provide insights on the relevance of the monomer binding pathway. The planed investigations will help to further detailed understanding of the structure-energetics relationships in DNA recognition by b-HLH-LZ proteins.

Appendix

Here I present some additional data collected in the course of the project. These experiments provide further information on the biophysical properties of Max variants and the Max/E-box complex. The primers used in mutation PCR to produce basic region X-to-Ala mutants are also given.

A. Determination of the van`t Hoff unfolding enthalpy of the core Max b-HLH-LZ domain Max$_{short}$.

CD melting experiments were performed in the standard working buffer with protein concentrations varying between 50 and 400 µM (monomer equivalents). MRE$_{222}$ was the experimental signal. The melting curves are shown in Figure A1(A).

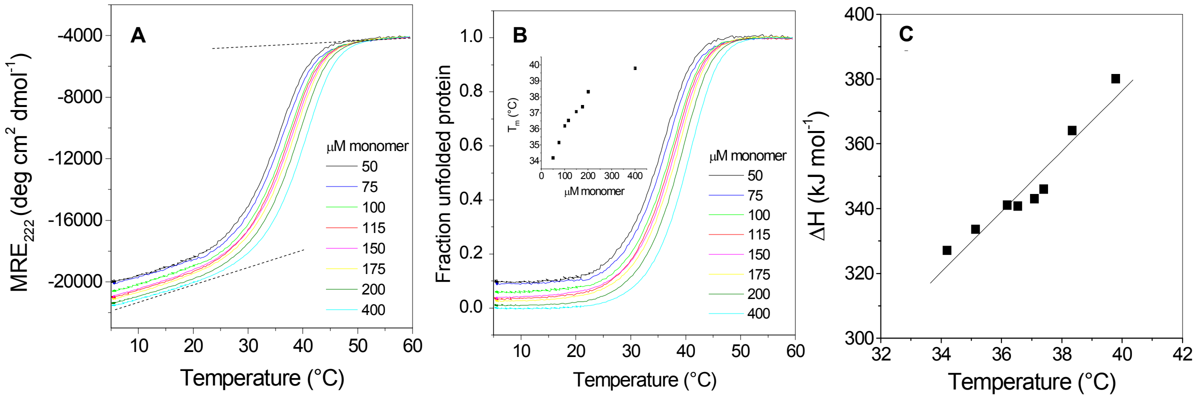

Figure A1. CD melting experiments with Max$_{short}$. The temperature tependence of MRE$_{222}$ and of the calculated fraction of unfolded protein are shown in A and B, respectively. The color code and the used concentrations are indicated. The insert in (B) shows the variation of T$_m$ with the concentration C,Temperature dependence of the unfolding enthalpy. See the text for details.

MRE$_{222}$ at 5 °C increases with the concentration, indicating equilibrium between folded dimer and unfolded monomer. Since MRE$_{222}$ of 200 and 400 µM solutions was almost identical, it was assumed that at above 200 µM the equilibrium is shifted to >99% to the dimeric state. The intrinsic temperature dependence of MRE$_{222}$ of the dimer and monomer, $\theta(T)^D$ and $\theta(T)^M$, respectively, are show with dashed lines in Figure A1(A). The fraction of unfolded protein was calculated as $f_U(T) = \frac{\theta(T) - \theta(T)^D}{\theta(T)^M - \theta(T)^D}$ (Figure A1(B)). The apparent van´t Hoff enthalpy was calculated from $\Delta H_{app} = 6RT_m^2\left(\frac{df_U}{dT}\right)$, where the derivative is taken in the temperature interval in which $0.4 > f_U > 0.6$. Since at lower concentrations $f_U \neq 0$, the true van´t Hoff enthalpy is $\Delta H = \Delta H_{app}/f_U$. As expected, ΔH increases with T (Figure A1(C)). In principle, the slope dΔH/dT yields the unfolding heat capacity, ΔC_p. From the slope of the regression line in Figure A1(C), $\Delta C_p = 9.3$ kJ M^{-1} mol^{-1}. This value is much higher than ΔC_p determined by DSC (2-3 kJ M^{-1} mol^{-1}; see main text). The reason is that T_m varies by only 5 deg between 50 and 400 μM concentrations. It is known that reliable ΔC_p values are calculated from Kirchhoff's plots if the T_m-variation is at least 20 deg. At the mean T_m in the set (36.8 °C), the mean ΔH is 345 ± 17 kJ mol^{-1}, in very good agreement with ΔH determined by DSC (see Table 3.3 in the main text).

B. Statistical thermodynamic analysis of the excess heat capacity of Max_{short}^{VL}

As discussed in section 3.1.4 in the main text, thermal unfolding of Max_{short}^{VL} follows a complicated mechanism, in which an unfolding, dimeric intermediate is significantly populated. The collected thermograms are presented in Figure B1. In this section I describe deconvolution of the heat capacity traces aimed at extracting quantitative information on the unfolding mechanism. At any temperature T, the heat capacity of the protein is defined as:

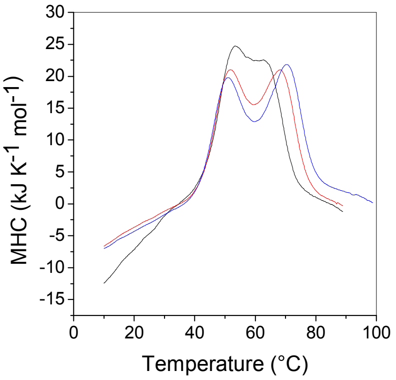

Figure B1. Heat capacity profiles of Max_{short}^{VL} collected with 30 μM (black), 78 μM (red) and 130 μM (blue) total monomer concentration

$$C_p(T) = C_{p,N}(T) + \frac{d\langle \Delta H \rangle}{dT} = C_{p,N}(T) + \sum_i f_i \Delta C_{p,i} + \sum_i \Delta H_i \frac{df_i}{dT} \tag{A1}$$

$C_{p,N}$ is the heat capacity of the native state, <ΔH> is the excess enthalpy over the enthalpy of the native state, f_i is the fractional population of each state i, and ΔH_i and $\Delta C_{p,i}$ are the enthalpy and the heat capacity change for the transition from state i-1 to state i. In the present case, since the solubility of Max_{short}^{VL} is limited and the used concentrations were relatively low, the pre-transitional and post-transitional heat capacities had erratic slopes (temperature dependence) and the first two terms on the right-hand side of the above equation could not be included in the analysis.

<ΔH> is linked to the system partition function (Q) by:

$$\langle \Delta H \rangle = RT^2 \frac{d \ln Q}{dT} \tag{A2}$$

The available biophysical information is compatible with the minimal unfolding model N → I → 2U, where the N → I and I → 2U transitions are characterized by the equilibrium constants K_{IN} and K_{UI}, respectively, and the fractional populations are f_N, F_I, and F_U. The system partition function for this model is:

$$Q = \frac{[N]+[I]+[U]}{N} = 1 + K_{IN} + K_{IN}K_{UI} \tag{A3}$$

and the population of states is $f_N = 1/Q$, $f_I = K_{IN}/Q$ and $f_U = K_{IN}K_{UI}/Q$. For the excess enthalpy and its temperature derivative we have:

$$\langle \Delta H \rangle = \Delta H_{IN} f_I + (\Delta H_{IN} + \Delta H_{UI}) f_U \tag{A4}$$

$$\frac{d\langle \Delta H \rangle}{dT} = \frac{d}{dT}(\Delta H_{IN} f_I) + \frac{d}{dT}\left[(\Delta H_{IN} + \Delta H_{UI}) f_U\right] = C_p(T) \tag{A5}$$

The explicit differentiation (which is not given here) leads to an expression combining ΔH_{IN}, ΔH_{UI}, K_{IN}, K_{UI} and Q, which are optimized by non-linear regression to minimize the function:

$$\chi 2 = \frac{1}{n^{eff} - p} \sum_i \left[\left(\frac{d\langle \Delta H \rangle}{dT} \right)_{i,calc} - \left(\frac{d\langle \Delta H \rangle}{dT} \right)_{i,experiment} \right]^2$$

n^{eff} is he number of experimental points and p is the number of adjustable parameters. The results of the modelling are shown in Figure B2. The correspondence between the modelled function and the experimental data is very good. Therefore, the assumed minimal model correctly captures the thermal unfolding mechanism of Max_{short}^{VL}. The calculation allows calculation of the fractional population of the native protein, unfolded protein and the dimeric intermediate. Figure B2 illustrates that unfolding starts around 50 °C in a concentration independent manner (The apparent shift of T_{m1} for 15 µM concentrated protein is most likely an artefact introduced by the very steep erratic increase of C_p at low temperatures; see Figure B1). However, the mid-point of the second transition is clearly depends on the concentration since it represents a dissociation event. As the consequence, the population of the intermediate becomes higher as the two transitions move away from each other with the variation of the concentration.

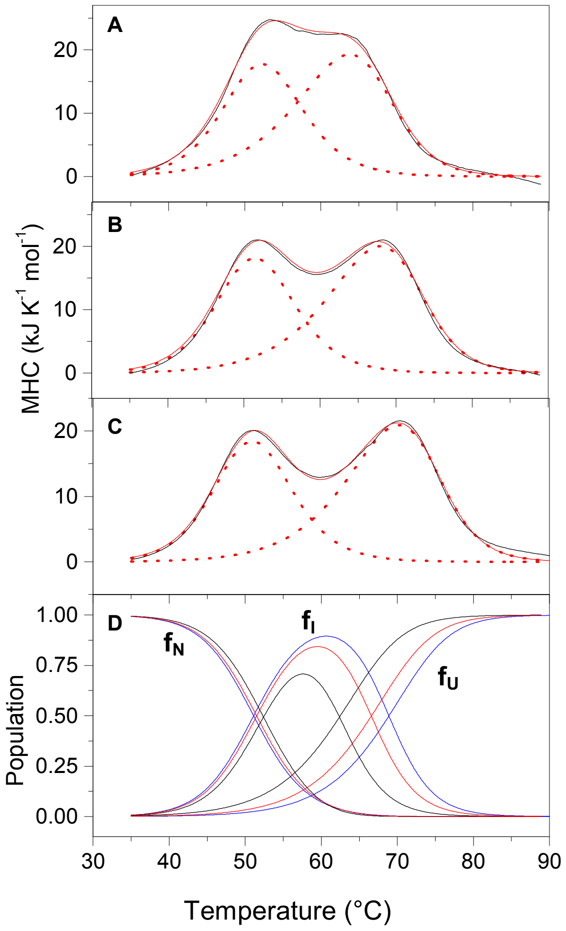

Figure B2. Simulations of the temperature dependence of the heat capacity of Max$_{short}^{VL}$. The protein concentration in monomer equivalents was 30 µM (A), 78 µM (B) and 130 µM (C). In panels A, B and C the experimentally measured C_p trace is shown in black. The best fit according to the combined equations A1-A5 is shown in red. The dotted lines are the simulated heat capacities of the two transitions. Note the pronounced broadening of the high-temperature transition at the pre-transitional side of the heat absorption peak. This is typical for unfolding linked to subunit dissociation. In contrast, the heat absorption peak of the monomolecular first transition is fully symmetric in respect to T_{m1}. transition In panel D, the calculated fractional population of the native state (N), the dimeric intermediate (I) and the denatured state (U) are shown. Black, 30 µM total monomer. Red, 78 µM monomer. Blue, 130 µM monomer. The parameters describing the two transitions are the following (concentration in µM monomer, T_m in units of °C, ΔH_m in units of kJ mol^{-1}):

Concentration	T_{m1}	ΔH_{m1}	T_{m2}	ΔH_{m2}
30	52.0	249	62.5	325
78	51.5	252	66.3	335
130	50.1	254	68.9	345

C. The heat capacity change of Max p21 and Max$_{short}^{SS}$ binding to E-box DNA

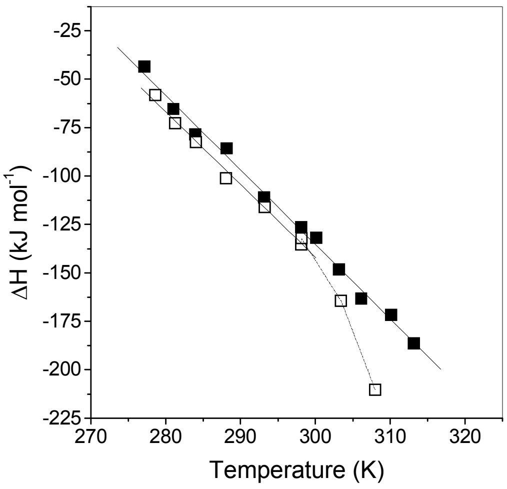

Figure C1

The high thermodynamic stability and the monomeric character of Max$_{short}^{SS}$ gives the opportunity to verify the heat capacity change, $\Delta C_{p,A}$, accompanying the formation of the protein-DNA complex. ITC binding experiments with Max$_{short}^{SS}$ were performed between 5 and 40 °C. As seen in Figure A2, ΔH changes linearly with T (filled symbols). The slope of the ΔH-vs-T plot defines $\Delta C_{p,A}$ with very good precision. The linearity suggests that the binding heat capacity is temperature independent. In contrast, Max p21 is much less stable and starts to dissociate to monomers at T>25 °C. Therefore, $\Delta C_{p,A}$ in this case is less well defined. However, the slopes of the

Kirchhoff's plots of Max p21 (open symbols) and $\mathrm{Max}_{\mathrm{short}}^{\mathrm{SS}}$ (filled symbols) are identical within error (±0.2 kJ K^{-1} mol^{-1}) between 5 and 25 °C. At T>25 °C a clear deviation from the linearity was observed in the Kirchhoff's plot of Max p21. The reason is that a substantial amount of Max p21 monomers are present in the concentration regime of the ITC experiments. Since DNA binding significantly stabilizes the protein, a shift of the monomer-dimer equilibrium takes place, and the apparent "excess" exothermicity is caused by the enthalpy of dimerization (folding). For that reason, binding experiments were performed between 5 and 25 °C.

D. Temperature-salt effects on the energetics of E-box binding by $\mathrm{Max}_{\mathrm{short}}^{\mathrm{SS}}$

Binding of $\mathrm{Max}_{\mathrm{short}}^{\mathrm{SS}}$ was studied by ITC at different salt concentrations at 5 °C and 25 °C. The binding enthalpy exhibits small variation with $[C^+]$, which is shown in Figure D1(A). The ΔH-vs-log$[C^+]$ plots at 5 and 25 °C are not parallel to each other, indicating a small heat capacity variation as function of the salt concentration (Figure D(B)). ΔC_p becomes more negative at high salt concentration, yet the difference between 0.12 and 1 M C^+ is roughly 0.3 kJ K^{-1} mol^{-1} and might be difficult to be detected experimentally, given that the effect is comparable with the typical uncertainty of ΔC_p estimates. Nevertheless, the observation is in line with the idea that the heat capacity of electrostatic interactions is slightly positive: At 1 M C^+ electrostatics is attenuated and dehydration of non-polar surface is the main contributor to negative ΔC_p.

Comparison of the salt dependence of binding affinities at 5 and 25 °C leads to the intriguing conclusion that at low temperature less counter cations are expelled from the protein-DNA complex (Figure D1(C)). The difference in the $\log K_A$-vs-$\log C^+$ slopes corresponds to 2 such cations (see section 3.2.4 in the main text for details). As far as the number of expelled cations equal the number of protein contacts to DNA backbone atoms, one can envisage temperature-dependent changes in the structure of the protein-DNA complex, leading to temperature-dependent formation/disruption of one non-specific contact per binding site.

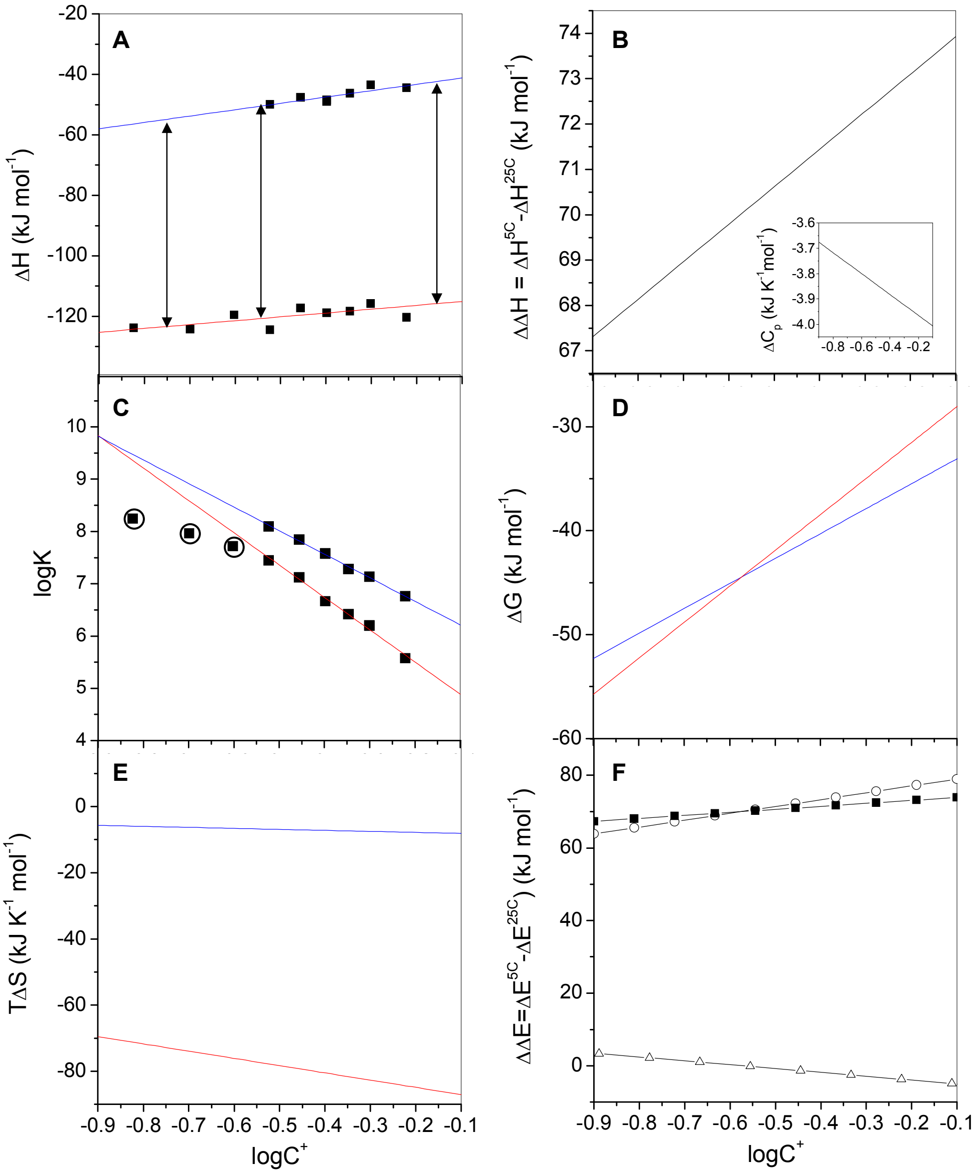

Figure D1. Temperature and salt effects on the energetics of DNA recognition by Max$_{short}^{SS}$. All plots are uniformly scaled in respect to logC$^+$. Data collected at 5 °C and 25 °C are shown in blue and red, respectively. A, Binding enthalpy. The large difference between 5 and 25 °C is caused by the negative binding ΔC_p. The arrows indicate the non-parallelity of the plots. The enthalpic difference increases at higher salt (B), leading to an increase in ΔCp of 0.3 kJ K^{-1} mol^{-1} between 0.1 and 1 M salt (inset in B). C, Binding affinity. The circled symbols illustrate that at low salt concentration where binding is strong, ITC data underestimate the genuine binding affinity. D, Stability of the protein-DNA complex. E, Total binding entropy. As in A, the large difference in the absolute entropies at 5 and 25 °C comes from the negative ΔC_p. F, Change in the energetic parameters of the binding reaction at 5 °C relative to 25 °C. Squares, ΔG. Circles, ΔH. Triangles, TΔS.

The difference in the salt dependencies of the binding constants decreases at lower salt concentration. K_A at 5 °C and 25 °C are the same at 0.13 M salt (Figure D1(C)). As the consequence, the protein-DNA complex is more stable at 5 °C than at 25 °C at relatively low salt conditions (0-0.26 M). At higher salt conditions, stability is higher at 25 °C (Figure D1(D)). Since the thermodynamic functions change linearly with temperature and the salt concentration in the nucleus is probably quite high (including the presence of >250 mM glutamate) one might speculate that at the physiologically relevant temperature of 37 °C the linked temperature-salt effects will add “extra” affinity in E-box recognition by Max.

The entropy of association is much more influenced by salt at 25 °C than at 5 ° (Figure D1(E)). In principle, this might indicate a lower entropic benefit from counter cation expulsion at lower temperature. However, the tested temperature range is probably too small to anticipate sizeable differences in the entropy of the bulk solvent and in the entropic characteristics of the counter cation cloud around DNA. Moreover, the entropic effect is probably dominated by general hydration differences and differences in the conformational entropy, since the basic region is significantly more folded (helical) at low temperature. Altogether, the presented results throw light on the complicated energy profile of protein-DNA complexes and might serve as a starting point of a systematic investigation of the combined effects of the temperature and salt on site-specific protein-DNA recognition.

E. Changes in helical content of Max p21 upon X-to-Ala mutations in the basic region

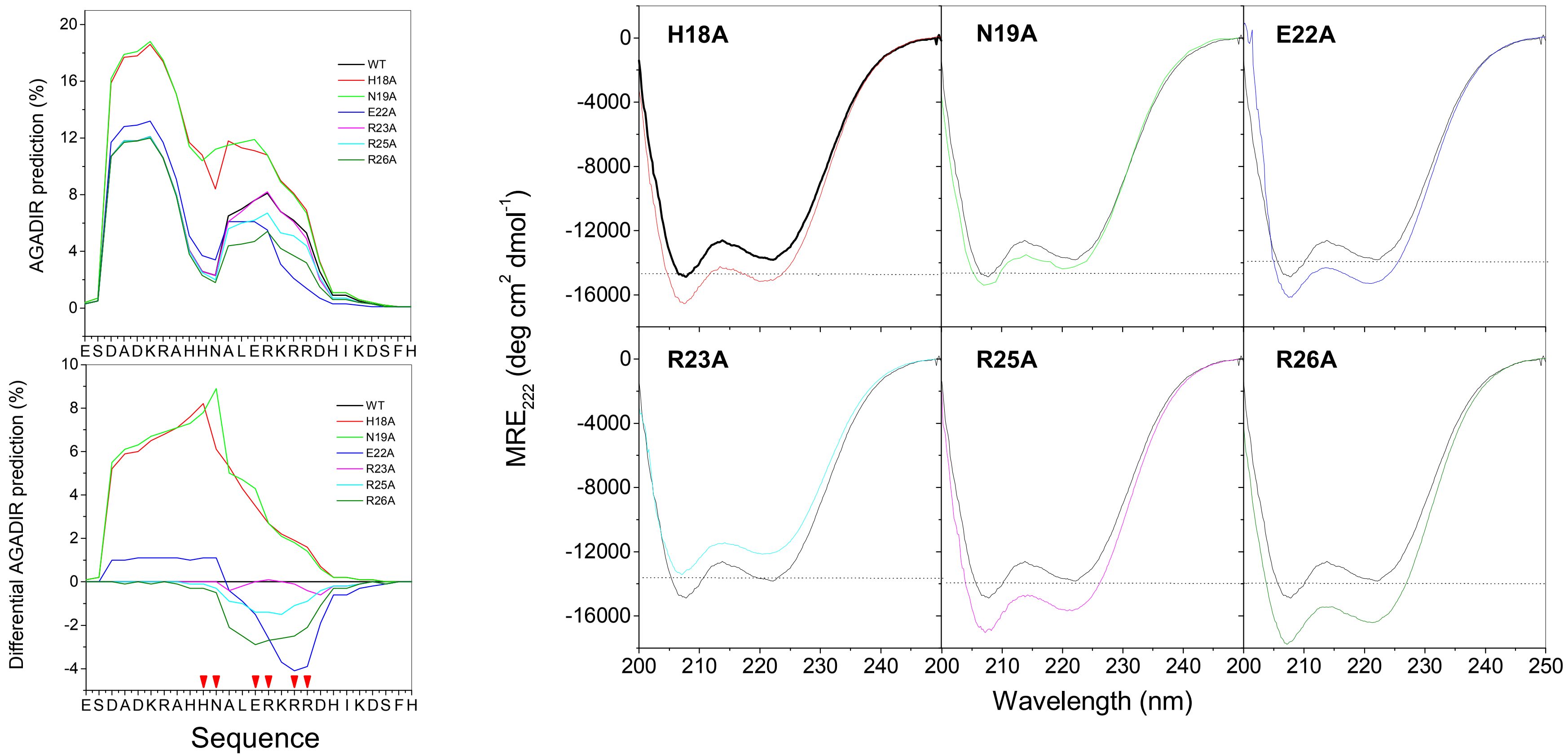

Figure E1. The predicted change of helical content upon X-to-Ala mutations was calculated with the algorithm AGADIR using the standard parameters at 25 °C and 0.2 M ionic strength (EMBL WWW gateway to AGADIR service). On the left, the results of the predictions are shown as absolute prediction (top panel) and differential prediction (mutant side chain– wild type side chain; bottom panel). The positions of the mutations are indicated with red symbols. As expected, the highest increase in helicity is expected for the H18A and N19A replacements, since both H and N have low helical propensities. Interestingly, the effect of these replacements (and of the E22A mutant) is long-range. The rest of the mutations are expected to have smaller and localized effects. Overall, AGADIR predicts small changes in helicity. On the right, the far-UV spectra of the 6 variants are shown. In each plot the spectrum of wild type Max p21 is drawn as the reference. The dashed lines represent the expected MRE$_{222}$, if the AGADIR prediction was correct. The experiment roughly confirms the prediction for H18A and N19A. The decrease in ellipticity is qualitatively confirmed also for R23A. In the rest three mutants, the mutation increases the helicity of the basic region, contra to the AGADIR prediction

F. Increase of the helical content of Max p21 and X-to-Ala mutants upon DNA binding

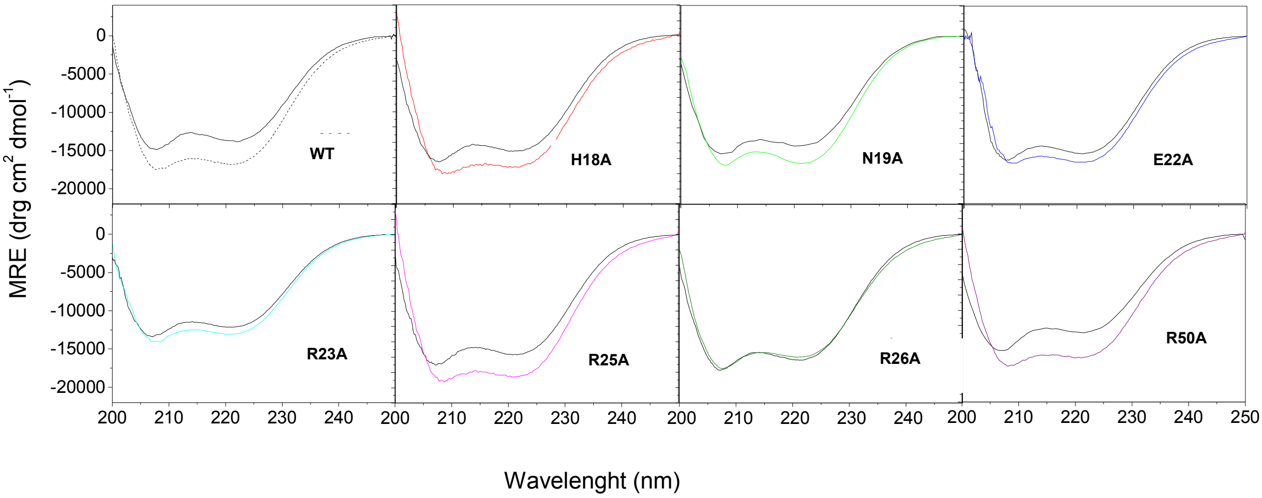

Figure F1. DNA binding is accompanied by transition of the basic region from a quasi-random-coil conformation to a well ordered α-helix within the DNA major groove. The figure compares the far-UV spectra of wild type Max p21 and the variant proteins in the free state (black lines) and in the DNA bound state (colored lines). In all experiments, identical concentrations of protein and 1:1 protein DNA complex were used. Depending on the binding affinity (K_A), the concentrations were selected such that the population of the protein-DNA complex was > 95% in all cases. (The latter condition doesn't gold for R26A, which binds in the high micromolar range.) In all cases but for R26A the ellipticity in the bound state is higher when compared with the free protein. However, the absolute increase in MRE_{222} vary from 900 to 4000 deg cm^2 dmol^{-1}, without any apparent correlation with neither the type of the replaced side chain, nor the type of contacts to DNA (base-specific hydrogen bonds or backbone contacts), nor the binding affinity, nor the intrinsic change in helical propensity introduced by the mutation. We note that the increase in helicity for side chains located in the middle of the DNA-recognition segment of the α-helix (E22 and R23) is the smallest. However, the significance of the observation is not obvious. It seems that there are local differences in the conformation of the basic region α-helix.

G. Oligunucleotides used as primers in PCR for site directed mutagenesis to produce X-to-Ala mutants in the basic region of Max p21.

Max p21 H18A

5’ GAC AAA CGG GCT CAT *GCA* AAT GCA CTG GAA CGA AAA CGT 3’

Max p21 N19A

5’ GAC AAA CGG GCT CAT CAT *GCA* GCA CTG GAA CGA AAA CGT 3’

Max p21 E22A

5’ CGG GCT CAT CAT AAT GCA CTG *GCA* CGA AAA CGT AGG GAC CAC 3’

Max p21 R23A

5’ CGG GCT CAT CAT AAT GCA CTG GAA *GCA* AAA CGT AGG GAC CAC 3’

Max p21 R25A

5’ CAT AAT GCA CTG GAA CGA AAA *GCA* AGG GAC CAC ATC AAA GAC 3’

Max p21 R26A

5’ GCA CTG GAA CGA AAA CGT *GCA* GAC CAC ATC AAA GAC AGC 3’

Max p21 R50A

5’ GGA GAG AGG *GCA* TCC GCG GCC CAA ATC CTA GAC 3’

LIST OF REFERENCES

1. Jelesarov, I., and Bosshard, H. R. (1999) Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition, *J. Mol. Recognit. 12*, 3-18.
2. Milev, S., Gorfe, A. A., Karshikoff, A., Clubb, R. T., Bosshard, H. R., and Jelesarov, I. (2003) Energetics of sequence-specific protein-DNA association: Conformational stability of the DNA binding domain of integrase Tn916 and its cognate DNA duplex, *Biochemistry 42*, 3492-3502.
3. Jen-Jacobson, L. (1997) Protein-DNA recognition complexes: Conservation of structure and binding energy in the transition state, *Biopolymers 44*, 153-180.
4. Cooper, A., Johnson, C. M., Lakey, J. H., and Nollmann, M. (2001) Heat does not come in different colours: entropy-enthalpy compensation, free energy windows, quantum confinement, pressure perturbation calorimetry, solvation and the multiple causes of heat capacity effects in biomolecular interactions, *Biophys. Chem. 93*, 215-230.
5. Dunitz, J. D. (1995) Win Some, Lose Some - Enthalpy-Entropy Compensation In Weak Intermolecular Interactions, *Chem. Biol. 2*, 709-712.
6. Sturtevant, J. M. (1977) Heat capacity and entropy changes in processes involving proteins, *Proc. Natl. Acad. Sci. U. S. A. 74*, 2236-2240.
7. Liu, L., Yang, C., and Guo, Q. X. (2000) A study on the enthalpy-entropy compensation in protein unfolding, *Biophys. Chem. 84*, 239-251.
8. Jayaram, B., and Jain, T. (2004) The role of water in protein-DNA recognition, *Annu. Rev. Biophys. Biomol. Struct. 33*, 343-361.
9. Schwabe, J. W. R. (1997) The role of water in protein DNA interactions, *Curr. Opin. Struct. Biol. 7*, 126-134.
10. Ellenberger, T. (1994) Getting a grip on DNA recognition: structures of the basic region leucine zipper, and the basic region helix-loop-helix DNA-binding domain, *Curr. Opin. Struct. Biol. 4*, 12-21.
11. Merabet, E., and Ackers, G. K. (1995) Calorimetric Analysis of Lambda-Ci Repressor Binding to DNA Operator Sites, *Biochemistry 34*, 8554-8563.
12. Luque, I., and Freire, E. (1998) Structure-based prediction of binding affinities and molecular design of peptide ligands, *Meth. Enzymol.. 295*, 100-127.

13. Milev, S., Gorfe, A. A., Karshikoff, A., Clubb, R. T., Bosshard, H. R., and Jelesarov, I. (2003) Energetics of sequence-specific Protein-DNA association: Binding of integrase Tn916 to its target DNA, *Biochemistry 42*, 3481-3491.

14. Ha, J. H., Spolar, R. S., and Record, M. T. (1989) Role of the Hydrophobic Effect in Stability of Site-Specific Protein-DNA Complexes, *J. Mol. Biol. 209*, 801-816.

15. Baker, B. M., and Murphy, K. P. (1998) Prediction of binding energetics from structure using empirical parameterization, *Meth. Enzymol., Pt B 295*, 294-315.

16. Cooper, A. (2000) Heat capacity of hydrogen-bonded networks: an alternative view of protein folding thermodynamics, *Biophys. Chem. 85*, 25-39.

17. Berger, C., Jelesarov, I., and Bosshard, H. R. (1996) Coupled folding and site-specific binding of the GCN4-bZIP transcription factor to the AP-1 and ATF/CREB DNA sites studied by microcalorimetry, *Biochemistry* 35, 14984-14991

18. Ladbury, J. E., Wright, J. G., Sturtevant, J. M., and Sigler, P. B. (1994) A Thermodynamic Study of the Trp Repressor-Operator Interaction, *J. Mol. Biol. 238*, 669-681.

19. Lundback, T., Chang, J. F., Phillips, K., Luisi, B., and Ladbury, J. E. (2000) Characterization of sequence-specific DNA binding by the transcription factor Oct-1, *Biochemistry 39*, 7570-7579.

20. O'Brien, R., DeDecker, B., Fleming, K. G., Sigler, P. B., and Ladbury, J. E. (1998) The effect of salt on the tata binding protein-DNA interaction from an hyperthermophilic archaeon, *Biophys. J. 74*, A241-A241.

21. Privalov, P. L., Jelesarov, I., Read, C. M., Dragan, A. I., and Crane-Robinson, C. (1999) The energetics of HMG box interactions with DNA: Thermodynamics of the DNA binding of the HMG box from mouse Sox-5, *J. Mol. Biol. 294*, 997-1013.

22. Huynen, M. A., and Bork, P. (1998) Measuring genome evolution, *Proc. Natl. Acad. Sci. U. S. A. 95*, 5849-5856.

23. Levine, M., and Tjian, R. (2003) Transcription regulation and animal diversity, *Nature 424*, 147-151.

24. Bergmann, D. J., Hooper, A. B., and Klotz, M. G. (2005) Structure and sequence conservation of hao cluster genes of autotrophic ammonia-oxidizing bacteria: Evidence for their evolutionary history, *Appl. Environ. Microbiol. 71*, 5371-5382.

25. Weiss, R. (1982) The Myc Oncogene in Man and Birds, *Nature 299*, 9-10.

26. Blick, M., Romero, P., Talpaz, M., Kurzrock, R., Shtalrid, M., Andersson, B., Trujillo, J., Beran, M., and Gutterman, J. (1987) Molecular Characteristics of Chronic Myelogenous Leukemia in Blast Crisis, *Cancer Genet. Cytogenet. 27*, 349-356.

27. Kurzrock, R., Kloetzer, W. S., Talpaz, M., Blick, M., Walters, R., Arlinghaus, R. B., and Gutterman, J. U. (1987) Identification of Molecular Variants of P210ber-Abl in Chronic Myelogenous Leukemia, *Blood 70*, 233-236.

28. Blackwood, E. M., Luscher, B., Kretzner, L., and Eisenman, R. N. (1991) The Myc Max Protein Complex and Cell-Growth Regulation, *Cold Spring Harb. Symp. Quant. Biol. 56*, 109-117.

29. Grandori, C., Cowley, S. M., James, L. P., and Eisenman, R. N. (2000) The Myc/Max/Mad network and the transcriptional control of cell behavior, *Annu. Rev. Cell Dev. Biol. 16*, 653-699.

30. Hurlin, P. J., and Huang, J. (2006) The MAX-interacting transcription factor network, *Semin. Cancer Biol. 16*, 265-274.

31. Kato GJ, W. D., Dang CV. (1992) DNA binding by the Myc oncoproteins, *Cancer Treat. Res.*, 313-325. .

32. Blackwood, E. M., and Eisenman, R. N. (1991) Max - a Helix-Loop-Helix Zipper Protein That Forms a Sequence-Specific DNA-Binding Complex with Myc, *Science 251*, 1211-1217.

33. Blackwood, E. M., Luscher, B., and Eisenman, R. N. (1992) Myc and Max Associate Invivo, *Genes Dev. 6*, 71-80.

34. Fernandez, P. C., Frank, S. R., Wang, L. Q., Schroeder, M., Liu, S. X., Greene, J., Cocito, A., and Amati, B. (2003) Genomic targets of the human c-Myc protein, *Genes Dev. 17*, 1115-1129.

35. Vennstrom, B., Sheiness, D., Zabielski, J., and Bishop, J. M. (1982) Isolation and Characterization of C-Myc, a Cellular Homolog of the Oncogene (V-Myc) of Avian Myelocytomatosis Virus Strain-29, *J. Virol. 42*, 773-779.

36. Thompson, E. B. (1998) The many roles of c-Myc in apoptosis, *Annu. Rev. Physiol. 60*, 575-600.

37. Tsuneoka. (2000) Ras/MEK signaling suppresses Myc-dependent apoptosis in cells transformed by c-myc and activated ras, *Oncogene 19*, Pages 115-123.

38. Ayer, D. E., Kretzner, L., and Eisenman, R. N. (1993) Mad - a Heterodimeric Partner for Max That Antagonizes Myc Transcriptional Activity, *Cell 72*, 211-222.

39. Hurlin, P. J., Queva, C., Koskinen, P. J., Steingrimsson, E., Ayer, D. E., Copeland, N. G., Jenkins, N. A., and Eisenman, R. N. (1995) Mad3 and Mad4 - Novel Max-Interacting Transcriptional Repressors That Suppress C-Myc Dependent Transformation and Are Expressed during Neural and Epidermal Differentiation, *EMBO J. 14*, 5646-5659.

40. Hooker, C. W., and Hurlin, P. J. (2006) Of Myc and Mnt, *J. Cell Sci. 119*, 208-216.

41. Hurlin, P. J., Queva, C., and Eisenman, R. N. (1997) Mnt, a novel Max-interacting protein is coexpressed with Myc in proliferating, rating cells and mediates repression at Myc binding sites, *Genes Dev. 11*, 44-58.

42. Hurlin, P. J., Steingrimsson, E., Copeland, N. G., Jenkins, N. A., and Eisenman, R. N. (2000) Mga, a dual-specificity transcription factor that interacts with Max and contains a T-domain DNA-binding motif *EMBO J. 19*, 3841-3841.

43. Makela, T. P. (1992) Alternative Forms of Max as Enhancers or Suppressors of Myc-Ras Cotransformation, *Science 256*, 373-377.

44. Krippner-Heidenreich, A., Talanian, R. V., Sekul, R., Kraft, R., Thole, H., Ottleben, H., and Luscher, B. (2001) Targeting of the transcription factor Max during apoptosis: phosphorylation-regulated cleavage by caspase-5 at an unusual glutamic acid residue in position P1, *Biochem. J. 358*, 705-715.

45. Pursglove, S. E., Fladvad, M., Bellanda, M., Moshref, A., Henriksson, M., Carey, J., and Sunnerhagen, M. (2004) Biophysical properties of regions flanking the bHLH-Zip motif in the p22 Max protein, *Biochem. Biophys. Res. Commun. 323*, 750-759.

46. Zhang, H., Fan, S. J., and Prochownik, E. V. (1997) Distinct roles for MAX protein isoforms in proliferation and apoptosis, *J. Biol. Chem. 272*, 17416-17424.

47. Brownlie, P., Ceska, T. A., Lamers, M., Romier, C., Stier, G., Teo, H., and Suck, D. (1997) The crystal structure of an intact human Max-DNA complex: New insights into mechanisms of transcriptional control, *Structure 5*, 509-520.

48. Nair, S. K., and Burley, S. K. (2003) X-ray structures of Myc-Max and Mad-Max recognizing DNA: Molecular bases of regulation by proto-oncogenic transcription factors, *Cell 112*, 193-205.

49. Anthony-Cahill, S. J., Benfield, P. A., Fairman, R., Wasserman, Z. R., Brenner, S. L., Stafford, W. F., Altenbach, C., Hubbell, W. L., and Degrado, W. F. (1992) Molecular Characterization of Helix-Loop-Helix Peptides, *Science 255*, 979-983.

50. Sauve, S., Tremblay, L., and Lavigne, P. (2004) The NMR solution structure of a mutant of the max b/HLH/LZ free of DNA: Insights into the specific and reversible DNA binding mechanism of dimeric transcription factors, *J. Mol. Biol. 342*, 813-832.

51. Naud, J. F., McDuff, F. O., Sauve, S., Montagne, M., Webb, B. A., Smith, S. P., Chabot, B., and Lavigne, P. (2005) Structural and thermodynamical characterization of the complete p21 gene product of Max, *Biochemistry 44*, 12746-12758.

52. Ferredamare, A. R., Prendergast, G. C., Ziff, E. B., and Burley, S. K. (1993) Recognition by Max of Its Cognate DNA through a Dimeric b/HLH/LZ Domain, *Nature 363*, 38-45.

53. Hu, J. Z., Banerjee, A., and Goss, D. J. (2005) Assembly of b/HLH/z proteins c-Myc, Max, and Mad1 with cognate DNA: Importance of protein-protein and protein-DNA interactions, *Biochemistry 44*, 11855-11863.

54. Cranz, S., Berger, C., Baici, A., Jelesarov, I., and Bosshard, H. R. (2004) Monomeric and dimeric bZIP transcription factor GCN4 bind at the same rate to their target DNA site, *Biochemistry 43*, 718-727.

55. Kim, B., and Little, J. W. (1992) Dimerization of a Specific DNA-Binding Protein on the DNA, *Science 255*, 203-206.

56. Kohler, J. J., Metallo, S. J., Schneider, T. L., and Schepartz, A. (1999) DNA specificity enhanced by sequential binding of protein monomers, *Proc. Natl. Acad. Sci. USA 96*, 11735-11739.

57. Kohler, J. J., and Schepartz, A. (2001) Kinetic studies of Fos center dot Jun center dot DNA complex formation: DNA binding prior to dimerization, *Biochemistry 40*, 130-142.

58. Wu, X. L., Spiro, C., Owen, W. G., and McMurray, C. T. (1998) cAMP response element-binding protein monomers cooperatively assemble to form dimers on DNA, *J. Biol. Chem. 273*, 20820-20827.

59. Marky, L. A., and Breslauer, K. J. (1987) Calculating Thermodynamic Data for Transitions of Any Molecularity from Equilibrium Melting Curves, *Biopolymers 26*, 1601-1620.

60. Pace, C. N., and Shaw, K. L. (2000) Linear extrapolation method of analyzing solvent denaturation curves, *Proteins 2000; Suppl.4:* 1-7.

61. Plotnikov, V. V., Brandts, J. M., Lin, L. N., and Brandts, J. F. (1997) A new ultrasensitive scanning calorimeter, *Anal. Biochem. 250*, 237-244.

62. Makhatadze, G. I., and Privalov, P. L. (1993) Contribution of Hydration to Protein-Folding Thermodynamics .1. The Enthalpy of Hydration, *J. Mol. Biol. 232*, 639-659.

63. Freire, E. (1995) Thermal denaturation methods in the study of protein folding, *Meth. Enzymol. 259*, 144-169.

64. Privalov, P. L., and Potekhin, S. A. (1986) Scanning microcalorimetry in studying temperature-induced changes in proteins, *Methods Enzymol. 131*, 4-51.

65. Wiseman, T., Williston, S., Brandts, J. F., and Lin, L. N. (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter, *Anal. Biochem. 179*, 131-137.

66. Karshikoff, A., and Ladenstein, R. (1998) Proteins from thermophilic and mesophilic organisms essentially do not differ in packing, *Protein Eng. 11*, 867-872.

67. Park, C., and Marqusee, S. (2004) Analysis of the stability of multimeric proteins by effective Delta G and effective m-values, *Protein Sci. 13*, 2553-2558.

68. Borisova, O. F., Shchyolkina, A. K., Chernov, B. K., and Tchurikov, N. A. (1993) Relative Stability of at and Gc Pairs in Parallel DNA Duplex Formed by a Natural Sequence, *FEBS Lett. 322*, 304-306.

69. Tchan, M. C., and Weiss, A. S. (2001) Asn(78) and His(81) form a destabilizing locus within the Max HLH-LZ homodimer, *FEBS Letters 509*, 177-180.

70. Manning. (1978) The molecular therory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides, *Q. Rev. Biophys. 11*, 179-246.

71. Plotnikov, V., Rochalski, A., Brandts, M., Brandts, J. F., Williston, S., Trasca, V., and Lin, L.-N. (2002) An autosampling differential scanning calorimeter instrument for studying molecular interactions., *Assay Drug Dev. Technol. 1*, 83-90.

72. Sauve, S., Tremblay, L., and Lavigne, P. (2004) The NMR solution structure of the max b/HLH/LZ free of DNA: Insights into the reversible DNA binding mechanism of dimeric transcription factors, *Protein Sci. 13*, 161-162.

73. Richardson, J. M., Lopez, M. M., and Makhatadze, G. I. (2005) Enthalpy of helix-coil transition: Missing link in rationalizing the thermodynamics of helix-forming

propensities of the amino acid residues, *Proc. Natl. Acad. Sci. U. S. A. 102*, 1413-1418.

74. Rose, G. D., Fleming, P. J., Banavar, J. R., and Maritan, A. (2006) A backbone-based theory of protein folding, *Proc. Natl. Acad. Sci. U. S. A. 103*, 16623-16633.
75. Dragan, A. I., Frank, L., Liu, Y. Y., Makeyeva, E. N., Crane-Robinson, C., and Privalov, P. L. (2004) Thermodynamic signature of GCN4-bZIP binding to DNA indicates the role of water in discriminating between the AP-1 and ATF/CREB sites, *J. Mol. Biol. 343*, 865-878.
76. Dragan, A. I., Liu, Y. Y., Makeyeva, E. N., and Privalov, P. L. (2004) DNA-binding domain of GCN4 induces bending of both the ATF/CREB and AP-1 binding sites of DNA, *Nucleic Acids Res. 32*, 5192-5197.
77. Olmsted, M. C., Bond, J. P., Anderson, C. F., and Record, M. T., Jr. (1995) Grand canonical Monte Carlo molecular and thermodynamic predictions of ion effects on binding of an oligocation (L8+) to the center of DNA oligomers, *Biophys. J. 68*, 634-647.
78. Bird, G. H., Lajmi, A. R., and Shin, J. A. (2002) Sequence-specific recognition of DNA by hydrophobic, alanine-rich mutants of the basic region/leucine zipper motif investigated by fluorescence anisotropy, *Biopolymers 65*, 10-20.
79. Milev, S., Bosshard, H. R., and Jelesarov, I. (2005) Enthalpic and entropic effects of salt and polyol osmolytes on site-specific protein-DNA association: The integrase Tn916-DNA complex, *Biochemistry 44*, 285-293.
80. Sharp, K. A. (1995) Polyelectrolyte electrostatics: Salt dependence, entropic, and enthalpic contributions to free energy in the nonlinear Poisson-Boltzmann model., *Biopolymers 36*, 227-243.
81. Sharp, K. A., Friedman, R. A., Misra, V., Hecht, J., and Honig, B. (1995) Salt effects on polyelectrolyte-ligand binding: comparison of Poisson-Boltzmann, and limiting law/counterion binding models, *Biopolymers 36*, 245-262.
82. Reddy, C. K., Das, A., and Jayaram, B. (2001) Do water molecules mediate protein-DNA recognition?, *J. Mol. Biol. 314*, 619-632.
83. Bergqvist, S., O'Brien, R., and Ladbury, J. E. (2001) Site-specific cation binding mediates TATA binding protein-DNA interaction from a hyperthermophilic archaeon, *Biochemistry 40*, 2419-2425.

84. Bergqvist, S., Williams, M. A., O'Brien, R., and Ladbury, J. E. (2004) Heat capacity effects of water molecules and ions at a protein-DNA interface, *J. Mol. Biol. 336*, 829-842.

85. Sauve, S., Naud, J. F., and Lavigne, P. (2007) The mechanism of discrimination between cognate and non-specific DNA by dimeric b/HLH/LZ transcription factors, *J. Mol. Biol. 365*, 1163-1175.

86. Griffith, K. L., and Wolf, J. R. E. (2002) A Comprehensive Alanine Scanning Mutagenesis of the Escherichia coli Transcriptional Activator SoxS: Identifying Amino Acids Important for DNA Binding and Transcription Activation, *J. Mol. Biol. 322*, 237-257.

ACKNOWLEDGMENT

First I would like to thank PD Dr. Ilian Jelezarov for giving me the opportunity to work on this interesting and challenging project. It was a pleasure. I would like to thank all the members of the Jelezarov group for their help, advices, fruitful discussions, encouragement and criticism. Especially I would like to thank Saša Bjelić for input, help and not to forget the discussions and coffee breaks.

I am also very greatful to Prof. Dr. Hans-Rudolf Bosshard and his group for their help, advices, input and fruitful discussions.

I would like to thank my husband Dr. Christian Meier for his help, his good advices, for encouraging me during hard times and for his patience. It was not always easy.

I am very grateful to my parents Béla and Eva Andrejszki for giving me the opportunity to study. Financially it was not always easy, thank you very much for your support.

Curriculum vitae

Personalien

Name	Meier-Andrejszki
Vorname	Laura Eva
Geburtsdatum	20. November 1974
Zivilstand	verheiratet
Nationalität	Doppelbürger Schweiz und Ungarn

Ausbildung

1991 - 1995: Handelsmittelschule, Kantonsschule Enge, Zürich

1995 - 1998: Eidgenössische Matura Typus E
AKAD, Zürich

1998 - 2002: Studium der Biochemie
Biozentrum der Universität Basel

2002: Diplomarbeit am Biochemischen Institut der Universität Zürich,
unter der Leitung von Prof. Dr. B. Gutte
"Mutants of the leucine zipper transcription factors c-Jun and c-Fos as HIV-enhancer binding peptides, inhibiting proviral transcription"

2003 - 2007: Doktorarbeit am Biochemischen Institut der Universität Zürich unter der Leitung von PD Dr. Ilian Jelezarov
“Energetics of E-box recognition by tumour suppressor Max p21”